Models to support eucalyptus plantations management under a changing environment

TESE APRESENTADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA FLORESTAL E DOS RECURSOS NATURAIS

Tânia Sofia Moreira de Paiva de Almeida Oliveira

ORIENTADORA: Doutora Maria Margarida Branco de Brito Tavares Tomé

JÚRI:

Presidente: Reitor da Universidade de Lisboa

Vogais: Doutor Daniel José Veja Nieva, Professor, Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, México

Doutora Maria Margarida Branco de Brito Tavares Tomé, Professora Catedrática, Instituto Superior de Agronomia, Universidade de Lisboa

Doutora Teresa de Jesus Fidalgo Fonseca, Professora Auxiliar, Escola de Ciências Agrárias e Veterinárias, Universidade de Trás-os-Montes e Alto Douro

Doutor Domingos Manuel Mendes Lopes, Professor Auxiliar, Escola de Ciências Agrárias e Veterinárias, Universidade de Trás-os-Montes e Alto Douro

Doutor Luís Mendes Godinho Milheiriço Fontes, Bolseiro de Investigação, Instituto Superior de Agronomia, Universidade de Lisboa

LISBOA
2015

Contents

Acknowledgements

I take the opportunity to thank the Portuguese Foundation for Science and Technology (FCT) that financed my PhD with grant SFRH / BD / 46594 / 2008.
Permission to use data from the ALTRI and PORTUCEL Portuguese pulp and paper companies is greatly acknowledged as well as the support from our colleagues Clara Araujo and Professor Carlos Arruda Pacheco in the characterization of the soil in the fertilization trials that allowed the model of the fertility ratio as a function of soil characteristics.

To all that helped me get here: thank you!

Lisbon, 31 of March 2014

Tânia Sofia Oliveira

Abstract

The objective of this work was the development of a tool to support eucalyptus stands management with the capacity of predicting forest development under different management options, thus providing forest managers with useful information in the form of variable outputs with interest for forest management. The tool uses the 3-PG model as the basis for growth predictions so that it can be used under a changing climate.

The first thing to be done was the improvement of the crown ratio equation, taking advantage of the great amount of available data. With more precise equations, new values of biomass where estimated and a new set of aboveground biomass equations was developed. Biomass values are not only an important model output, but also a vital piece in the hybridization of the GLOBULUS growth and yield model and the 3-PG whole stand process based model. The linkage of the models resulted in a hybrid model with more detailed outputs that were further complemented with a diameter distribution model.

The 3-PG fertility ratio is an important parameter of the model, but is estimated in a subjective way. The improvement with an equation that predicts it from soil characteristics is important not also because it is a flaw that is recurrently appointed to the model, but also because it will allow it to be sensible to fertilizations.

Key-words: forest management support, climate change, *Eucalyptus globulus*, process based models, 3-PG.

Resumo

O objectivo deste trabalho foi produzir uma ferramenta de apoio à gestão de povoamentos de eucalipto capaz de prever o desenvolvimento da floresta sob diferentes opções de gestão, fornecendo aos gestores informação útil sob a forma de outputs com interesse para a gestão florestal. A ferramenta usa o modelo 3-PG como bases das predições de crescimento para poder ser usado em condições de clima variáveis.

O primeiro passo foi o desenvolvimento de novas equações de proporção e profundidade da copa com vista à melhoria das estimativas de biomassa aérea por componentes. Isto permitiu não só o ajustamento de um sistema de equações ao nível do povoamento para as diferentes componentes da biomassa aérea como permitiu também constituir a base de trabalho para a hibridação do modelo empírico GLOBULUS e do modelo de base processual 3-PG. A ligação destes modelos permite obter variáveis de output mais detalhadas que foram posteriormente complementados com um modelo de distribuição de diâmetros.

O parâmetro de fertilidade do 3-PG é um parâmetro importante do modelo, mas é estimado de forma subjectiva. O melhoramento com uma equação que o estime a partir de características do solo é importante não só por esta ser uma falha recorrentemente apontada ao modelo, mas também porque vai permitir que o mesmo seja sensível a fertilizações.

Palavras-chave: apoio à gestão florestal, alterações climáticas, *Eucalyptus globulus*, modelos de base processual, 3-PG.

Thesis structure

The thesis is presented in the form of articles. Each article will have its goals, methods and conclusions, but all contributing to the overall objective of the thesis, which is organized as follows:

Introduction – in this chapter, the motivation and objectives of the work will be presented, alongside with the framework for each of the papers in the thesis structure and state of the art.

Chapter 1 – A system of compatible models to predict and project tree crown ratio and crown length for eucalypt plantations

Chapter 2 - Modelling the aboveground stand biomass of planted and coppiced *Eucalyptus globulus* in Portugal

Chapter 3 - Prediction of diameter distribution in whole stand process based models. An application to 3-PG calibrated for eucalyptus stands in Portugal

Chapter 4 - Modelling the 3-PG fertility ratio as a function of available soil water and other soil parameters

Chapter 5 - Using stand level allometric equations to hybridize process-based and empirical forest growth models. An application to 3-PG calibrated for eucalypt stands in Portugal

Final remarks – summarizes the main results of this work and points out future research need

Introduction

I.1. Background

Eucalyptus globulus Labill., an Australian native tree, was introduced in Portugal around 1850. Initially with only ornamental interest, it soon reached the attention of foresters, not only for the characteristics of its wood, but also due to its rapid growth in a large majority of the country. However, it was not until 1926, with the construction of the first Portuguese - and presumably also the worlds - factory which used eucalyptus as raw material for pulp, that the species began to gain real importance. Since then, we have witnessed a more or less parallel growth of the pulp industry and plantation area, the latter growing from about 70,000 ha in the 1960 (DGFSA 1966a, 1966b) up to about 740,000 ha of pure and mixed stands dominated by eucalyptus currently available in the country (AFN, 2010).

It is a fast growing species, mainly used these days by the pulp industry. The trees are planted at final density, as thinning and pruning practices are not usual during the first cutting cycle. The stands are intensively managed as a short rotation coppice system in which the first cycle of planted seedlings (single stem) is followed by 2 or 3 coppiced stands, with an average cutting cycle of 10–12 years (Soares and Tomé, 2001). Pulp and paper industries alone hold the management of about 154,000 ha of eucalyptus stands (CELPA 2011) and the remaining area is privately owned. Forest management of this species is an important issue, since the national production can no longer respond to the industry´s demands, which means that the industry has to rely on wood importation. In 2012, about 23% of the national consumption of the eucalyptus industry was satisfied with imported wood. The production increase can be accomplished through an increase in the forest productivity, not only using genetic improvement, but also improving the silvicultural treatments (Santos *et al.*, 2013).

The global context of today´s world, characterized by constant changes in society demands, increasing needs, soil use changes and, maybe most important than all of that, climate change, have been demanding an extra effort from forest management to adapt in order to accomplish this new demands in a sustainable way. Besides all this pressure, there has been a demand, from the forest management side, of simple, easy to use models that give more complete answers, i.e., that supply a whole group of variables indispensable for forest management and analysis of its sustainability. Until recently, only empirical growth and yield models were available in Portugal for *Eucalyptus globulus*. Several models were developed (Amaro, 2003; Amaro *et al.*, 1998; Barreiro *et al.*, 2004; Soares and Tomé, 2003; Tomé *et al.*, 2001) prior to the joint effort of University and industry to develop a model that could be used in the whole country, the GLOBULUS model (Tomé and Ribeiro, 2000; Tomé *et al.*, 2001; Tomé *et al.*,

2006). The GlobLand project improved the existing GLOBULUS 2.1 (Tomé *et al.*, 2001) and developed GLOBULUS 3.0 (Tomé *et al.*, 2006), the model that is currently in use. It is a stand level growth and yield model developed for pure even-aged stands that integrates all the information available on eucalyptus growth and yield in Portugal, representing the combined efforts between industry and universities, which have been involved in several co-operative research projects over the past decades. GLOBULUS 3.0 incorporates climatic information in the equations, but it is still an empirical model that lacks the flexibility and capacity to simulate environmental stresses and conditions beyond those of the data used to develop it.

Over the last years, process-based models have been gaining importance in forest management and in eucalyptus stands in Portugal. For eucalyptus plantations, the 3-PG (Landsberg and Waring, 1997; Sands and Landsberg, 2002) model is becoming more and more relevant. This model was developed as a simple process-based model that uses readily available data as input and it has already been parameterised for *E. globulus* in Portugal (Fontes *et al.*, 2006), but there is still some room for improvements, namely the lack of forest management interest variables in the outputs and the empirical way of estimating one of the key parameters of the model that deals with soil fertility, a key issue in *E. globulus* forest management in Portugal.

I.2. State of the art

Forest management has been undergoing an evolution that demands that multiple objectives are addressed aside from the traditional maintenance of wood resources levels. To meet this objective, a wide variety of management techniques and silvicultural systems have been developed, the nature and complexity of which have expanded dramatically in the last 20 years. The traditional even-aged, single-species plantation management is evolving to multi-objective forest management that requires a wide variety of management systems that are developed and selected based on ecosystem characteristics and management objectives. It has been widely accepted from the end of the XX century that although relatively simple growth and yield models remain fundamental for forest management, the global community now demands more sustainable use of multiple values and a more comprehensive understanding of forests and forest products (Vanclay, 1994). However, not much has been accomplished in terms of tools available to support forest managers in this new paradigm of forest management.

Models are simplifications of the real systems and are used when it is easier to work with the model than with the real system (Ford, 1999). Forests are complex ecological systems and forest management is a complex social, political, economic and biophysical activity. Multi-value sustainable forest management involves complex planning and the application of knowledge

from several sciences, which makes it one of the most complex human activities. At a basic level, modelling is the simple process of thinking systematically about a problem (Jakeman *et al.*, 2008) and modelling in forest management involves organization of data, assumptions and knowledge for a specific purpose. Modelling methodologies continue to evolve as new knowledge, powerful computer technology and advanced statistical techniques are brought to help growth and yield predicting. Levins (1966) argued that modellers of population biology strive to maximize simultaneously three desirable properties of a given model: generality, reality and precision. Generality refers to the applicability to a range of instances; reality might be thought as the conformity of model assumptions and relationships to the real system; precision indicates the degree of exactness in predictions. In any model, Levins asserted, developers may sacrifice one of these desired properties to achieve a higher level of the others. In traditional growth and yield models, generality is sacrificed for increased reality and precision, being the primary focus on precision, something that can be considered a reasonable strategy. On the other hand, in today's world of rapidly changing conditions, the interest in the generality property of the models has been increasing.

Several authors have classified forest growth models in 3 main types: empirical, process-based or mechanistic and hybrid. In some cases, process-based and hybrid models overlap, as the differences between them are not always clearly defined.
Empirical or statistical models use large data bases of measurements and statistical techniques like regression to derive quantitative relationships between growth and several tree and/or stand variables and have no basis on physiological process. They are important for updating forest inventories, comparing forest silvicultural treatments and estimating sustainable levels of harvest. Empirical models are robust, have a long story of development, rely on data that is generally available, the output is geared for operational decisions and can represent a wide range of conditions and sampling schemes. However, the data required has to be of high quality, the extrapolation power is very low and are generally insensitive to climate (Taylor *et al.* 2009). Since they are site-specific, no new variations in management or in environmental conditions can be considered, which makes it almost impossible for the models to be used outside the area and conditions for which they were developed. They are implemented in tools very commonly used in management and planning, because the information on forest productivity is accurate and reliable (Levy *et al.*, 2004) even if the use is limited to forest management practices for which there are data available (Garcia-Gonzalo *et al.*, 2007).

A process-based model (PBM) can be defined as a description of the behaviour of a system in terms of a set of functional relationships and their interactions with each other and the system environment. The term mechanistic, also applied to these models, is related to the fact

that the behaviour of the entire system is simulated in terms of a series of sub-models describing the processes/mechanisms that determine that behaviour and the responses of those processes to changes in environmental inputs. Process-based models represent key plant physiological processes like photosynthesis, which are then scaled to the stand-level to estimate growth. They are developed to understand the underlying mechanisms influencing growth, to test hypothesis about plant behaviour and to predict potential forest productivity. The mechanistic models can theoretically extrapolate to novel situations and are sensitive to climate, but depend on several difficult to measure parameters, the input data is not widely available, they have high computational demand and often the output is not enough for operational decisions (Mäkelä et al, 2000; Landsberg, 2003). The driving variables – the independent factors that determine the rates of particular processes – are related to environment. Photosynthesis, often regarded as the primary plant process, is driven by radiant energy; respiration rate is affected by temperature; the rate of nutrient supply to root systems is affected by water movement to the roots and by rates of nutrient diffusion. The change of state of plants with time in terms of these factors may be formally stated as differential equations.

PBMs are useful for long term predictions, especially with climate and silvicultural changes. Landsberg (1986) describes PBMs as mathematical representations of biological systems that incorporate knowledge about physiological and ecological mechanism into prediction algorithms and fundamental scientific tools that enable hypotheses formalization and a structure that synthesizes the available knowledge. However, PBMs are seldom used as tools in forest management because it is considered that they include too many uncertainties and require parameters that are not well known and sometimes difficult to get so that their projections are as reliable in practice as the one´s from empirical models (Mohren and Burkhart, 1994). Furthermore, there is an idea that the conventional approach to statistical modelling of the growth and productivity is far superior.

Empirical and process-based models were initially considered as mutually exclusive, although Korzukhin *et al.* (1996) stated that there are no pure empirical or PBM, being all models developed using both methodologies. There is currently a great interest in the application of PBMs in forest management, namely the inclusion of process-based elements in management models to make better use of empirical observations (Sharpe, 1990). There are already some experiences of this type of models, including stand level models. Most of PBMs are based on photosynthesis, considering it either as the basic growth process (Bartelink, 2000; Ditzer *et al.*, 2000; Lindner, 2000; Mäkelä *et al.*, 2000a; Raulier *et al.*, 2000; Valentine *et al.*, 2000) or using it as an independent prediction variable (Brunner, 2000; Courbaud, 2000). Other physiological processes, such as water balance and nutrient cycles have received less attention in the

management context (Ågren, 1996; Landsberg and Waring, 1997; Sands *et al.*, 2000), although they are fundamental for the perception of photosynthesis control and the effects of climate change on tree growth (Thornley and Cannell, 1996; Kirschbaum, 2000; Lindner, 2000) as well as regional variations in productivity (Landsberg and Waring, 1997; Coops *et al.*, 1998; Sands *et al.*, 2000).

PBMs are written in terms of mechanisms that rule the growth and respond to environmental changes, so, with the correct parameterization, they can be used in situations outside the ones for which empirical data was available, so they are usually used to answer the "why" or "how" questions, rather them the "what is" or "what if" ones (Leary, 1985). Korzuhkin*et al.* (1996) compared empirical and PBMs and concluded that PBMs offer significant advantages in forest management. Johnsen *et al.* (2001) discussed the use of PBMs as tools in research and also in forest management and favoured their use, noting however that its use in forest managing requires some simplifications.

Some important drawbacks to PBMs are that they are often quite complicated with outputs that are of little practical interest and can be difficult to parameterize or even operate due to high input data requirements. Matala *et al.* (2003) compared a PBM (FinnFor) and a statistical model (Motti) and found out that both gave similar predictions of relative growth rate, but the statistical model was more stable in terms of its predictions and not as sensitive to initial stand conditions and silvicultural treatments. Schmid *et al.* (2006) suggested that PBMs be preferred for simulations up to 100 years, because they account for changing climatic conditions. On the other hand, they found PBMs to be quite sensitive to uncertainties in model structure, parameter values and input data, which can be difficult to handle in large-scale applications (Schmid *et al.*, 2006). So, the key limitations of PBMs are initialization, parameterization, scale and sensitivity. Use of PBMs in an applied forest management context is limited, but has been increasing in recent years due to the new series of challenging questions that forest management faces (Mäkelä *et al.*, 2000a).

The state of the art in PBMs and their applicability to forest management can be summarized in three points:

1) The carbon balance is a sound basis for forest management applications and is based on estimates of photosynthetic production that is used to derive the growth of trees and stands.

2) The practical implementation of PBMs may be speeded when it is accepted that both models, empirical and process-based, can be improved with this integration.

3) The incorporation of PBM´s into forest management systems requires cooperation between modellers and forest managers (Mäkelä *et al.* 2000).

Process-based models have a long history in forestry. Several conceptual models of forest productivity were developed in the early 1980´s, but their scope of application was limited (Landsberg, 2003). Although there is good general knowledge of the physiological factors that influence forest production, the development of forest PBMs has been limited due to the lack of knowledge on key tree growth processes, especially the ones occurring below ground; trees are large, complex, multigenic, perennial organisms, and basic research is difficult; computers with enough processing power only recently have become available (Isebrands *et al.*, 1990). The use and application of PBMs has increased in the last years and several technical reviews have been presented (Ryan *et al.*, 1996; Battaglia and Sands, 1998; Mäkelä *et al.*, 2000a; Landsberg, 2003). The use of PBMs in research has increased in recent years not only because of the increased knowledge of biology and forest ecology together with the great technological advances in computers and software (Landsberg *et al*., 1991), but also due to the need to integrate more complex organic information to answer questions that are at higher scales than those on which the processes are measured (Smith *et al.*, 1998). This integration has been made to assist the interpretation of research activity, to address policy issues and to develop forest management tools. A wide range of issues are being considered, namely: how climate change will affect productivity (McMurtrie and Wang, 1993; Valentine *et al.*, 1997)? How can forest management be improved to increase the productivity of a stand (Sievanen and Burk, 1992)? How to account for the impact of intensive forestry practices on forested watersheds hydrology (Amatya and Skaggs, 2001)?

It is currently not possible to find a model that satisfies all the requirements for assessing the sustainability of a multi-functional forest management, i.e., a model that: uses the available data, such as forest inventories, as input; takes into account the genetics of the plant material; gives good predictions in situations of climate and forest management change and provides information not only in terms of wood production, but also of other forest products and services. A major challenge is the development of methodologies to combine the two types of models in order to gain the benefits from the advantages of both, and in recent years several authors have tried this process, usually known as the hybridization of empirical and process models (Baldwin *et al.*, 2001; Battaglia *et al.*, 1999; Almeida *et al.*, 2003; Robinson and Ek, 2003).

Hybrid models attempt to combine the strengths of both process-based and statistical models and are often described as the future of forest growth modelling (Landsberg, 2003). They offer several advantages that make them appealing to a wide audience (Taylor *et al.*, 2009). The primary advantages of hybrid models are that they: increase the flexibility of statistical models to address variability of climate and regional differences; simplify required inputs associated with PBMs and mechanistically represent influences of silvicultural treatments such as thinnings and fertilization. But there are also some disadvantages, they can compound error,

they still require difficult to access data and they bring generally small gain in accuracy when compared to statistical models. There is still a small number of hybrid models in the literature compared to the number of statistical models or PBMs, but their number has been increasing, especially after the advent of the 3-PG framework offered by Landsberg and Waring (1997).

Hybrid models are simply the combination of statistical and mechanistic approaches, often with an applied goal (Johnsen *et al.*, 2002). The main objective is to use physiological principles as much as possible while relying on allometrics and other accepted statistical modelling conventions for the rest of the attributes of interest. The scale used is mostly a stand-level and monthly basis, but there are tree-level and daily basis models too. The objectives of these models vary from predicting effects of climate change on tree growth (Baldwin *et al.*, 2001) to the role of silviculture on external log characteristics (Mäkelä and Makinen, 2003). Hybrid models combine statistical and process approaches in an attempt to take advantage of the strengths of both approaches. They are useful for predicting growth using climatic factors and novel forest silvicultural treatments. They are robust, sensitive to climate, can use traditional forest inventory data and minimize the number of required parameters, but the climate and soils input data may not be widely available and the accuracy improvements can be minimal when compared to a purely statistical approach (Monserud, 2003).

Robinson and Ek (2003) point out that the use of established sub-models already identified as important for forest growth and dynamics saves a lot of time. Doing this allows for much of the excellent work already made to be reused and that time is not spent in new sub-models development and programing each time a new model is developed. Asides from that, taking models/sub-models out of their context and using them in new frameworks allows the identification of short-comings that might have been missed in the original context. They concluded that it is reasonable to extract components from different forest growth and dynamics modelling traditions, combine them in a new way and expect to be able to make defensible predictions. Most of the key processes of forest growth have already been modelled in very precise ways and there is no justification for extra work. For example, light interception, photosynthesis and calculation of Gross Primary Production are areas where a lot of work has been done and there are several models ranging from very simple to more complex ones.

Several authors (e.g. Korzukin *et al.*, 1996; Johnsen *et al.*, 2001; Landsberg, 2003b) suggest that empirical or statistical models developed from historical observations of forest growth with little representation of ecological processes have limited application as decision-support tools to guide the application of ecosystem-based management. The limitations of such models are even more pronounced when the long-term impact of climate change arte taken into consideration. Process-based models are much more flexible and allow the simulation of changing environmental conditions, but their application as decision-support tools in forest

management has been limited for the difficulty to calibrate very complex models. Hybrid models have been developed in order to capture the flexibility of process-based simulations while maintaining the reliability of the historically based empirical models. Hybrid forest growth models use the knowledge of the growth process in a way that allows the integration of standard forest mensuration inputs and outputs. The link to the mensuration inputs and outputs is often achieved through statistical relationships (Mäkelä *et al.*, 2000). Several authors have attempted a classification of stand-level forest growth models (Korzukhin *et al.*, 1996; Landsberg, 2003b), but Mäkelä (2009) presented a very useful scheme to distinguish the three main hybrid modelling approaches, focusing the method of the hybridization and recognizing the core simulation approach used in the model: hybridized empirical models, hybridized process models and reduced-form process models.

Hybridized empirical models use process elements as sub-models to provide additional inputs, often expressed as modifiers of statistically derived growth functions, to empirical growth models. The sub-models can be simulations of intercepted radiation, gross or net photosynthetic production, soil water status, climate relationships or physical constraints such as growing space. This type of models is typically based on statistical relationships between height and diameter growth and site or environmental variables. It requires large databases to derive the relationships and the resulting models tend to be very site-specific (relationships based on correlations don´t always capture the interactions between process). One of the recognized strengths of this approach is the representation of competition between stems in a stand. Examples of this type of models are SORTIE-BC (Canham *et al.*, 1999) and FVS-BGC (Milner *et al.*, 2003).

Hybridized process models are the ones where potential growth rates, biomass allocation patterns and mortality functions are derived from empirical data and then modified using a complex system of physiological based functions or sub-models. The efficiency of these models depends on how the processes capture the key factors that regulate the growth rates. Examples of this type of models include the FORECAST model (Kimmins *et al.*, 1999) and many of the gap models derived from the work of Botkin *et al.* (1972) and Botkin (1993). Mäkelä (2009) suggests that this type of models have a good potential to capture the combined strengths of biometrics and process modellers.

The reduced-form process models represent a distinct subset of hybridized process models derived from detailed mechanistic models through an aggregation of growth processes and parameters into core components in order to increase flexibility and make them more user-friendly. The empirical elements not only relate to processes, the allocation of carbon within and between tress, for example, but also to the process of model parameterization and calibration. One of the benefits of this type of models is that the mechanistic foundation

provides insight on the impact of environmental factors on the productivity of different species. Model parameters can be estimated using different approaches, from Monte Carlo techniques (e.g. Mäkelä, 1988) to Bayesian synthesis (e.g. Green *et al.*, 1999; Van Oijen *et al.*, 2005). This type of models relies on empirical sub-models that translate productivity into mensurational output such as height and diameter increments. Examples of this type of models are 3-PG (Lansdberg and Waring, 1997) and Pipestem (Valentine *et al.*, 1997).

3-PG (Physiological Principles in Predicting Growth) is a simple process-based model, developed by Landsberg and Waring (1997) in order to bridge the gap between somehow simplistic empirical models and the more complex models with a physiological basis. The idea was to develop a practical tool, accessible to both scientists and forest managers. Following this idea, simplified description of the biological processes where used and combined with empirical relationships. It is a simple forest growth model developed with the specific aim of obtaining detailed information at the stand level and relevant to forest management (annual change in volume, diameters and biomass, among others). There is a strong correlation between the outputs of the model and empirical measurements (Landsberg *et al.*, 2001, 2003; Law *et al.*, 2001), consistent with forest inventory data obtained periodically during the rotation (Landsberg *et al.*, 2003; Waring and McDowell, 2002). The model has been well received and it is widely used, with currently over 100 peer-reviewed publications. It is based on physiological processes that were considered the most important, it requires a number of parameters and initialization data relatively low and easy to get compared to other PBMs and is implemented in Visual Basic for applications (in Excel environment), which makes it user friendly. 3-PG is available free on the internet, it has been tested and implemented in many countries (different soil and climatic conditions) and various species - *Eucalyptus globulus* (Sands and Landsberg, 2002; Fontes *et al.*, 2006), *Eucalyptus grandis* (Almeida *et al.*, 2004; Esprey *et al.*, 2004), *Pinus taeda* L. (Landsberg *et al.*, 2001), *Picea sitchensis* (Bong.) Car. (Waring, 2000), among others. All these features make it an excellent choice and ideal for this doctoral work. It is a generalized growth model (i.e., not specific to each site, but that needs to be parameterized for individual species) for even-aged forest stands. It requires few initialization variables to characterize the site and easy to obtain climatic variables. It foresees the development of a stand, the use and available water in the soil over time. It gives as primary response variables the net primary productivity, carbon allocation to leaves, stem and roots, number of trees per hectare, the available soil water and transpiration on a monthly basis. It is a model that deals primarily with the biology and physiology of the growth process, but it has a module that converts the biomass pools (a biological output) into variables with forest management interest, such as leaf area index, basal area, quadratic mean diameter, volume and mean annual increment in volume. The monthly time step allows the model to reflect the

effects of factors that would be masked in yearly time steps, such as water stress and temperature variations, but a month can also be too long a time step for more detailed information such as radiation interception and canopy photosynthesis. The nature of the data used in the model is then determined by the objectives. The input data are divided in three classes: weather data, site-specific factors and stand initialization data:

- Climatic data: monthly mean values of solar radiation, monthly averages of minimum and maximum air temperature, monthly total precipitation and number of days per month with rain and with frost. These values can be the ones that were observed or, if not available, historical data, long-terms averages or scenarios for future climate.

- Site-specific factors: latitude, site fertility rate and the maximum value of water the soil can hold.

- Stand initialization data: initial values for the number of trees per hectare, biomass of foliage, stem and roots and the available soil water in the root zone.

The primary outputs from the model are the state variables: biomass (wood + bark + branches, leaves and roots), the number of trees per hectare and the available water in the soil, asides from some stand-level information that is more relevant for forest management, such as stand basal area, the average diameter, the volume under bark and the mean annual increment in volume. There are also other variables of interest, such as gross primary production, net primary production, stand evapo-transpiration and canopy leaf area index that are determined as part of the updating the state variables.
It is possible to identify five sub-models in 3-PG: the assimilation of carbohydrates, the carbon allocation to leaves, stem and roots, the evolution of the number of trees per hectare, the water balance in the soil and the conversion of biomass into variables with interest for the management (Sands and Landsberg, 2002):

1. Carbohydrates assimilation (net and gross primary productivity)

Calculation of gross primary productivity (GPP) is made using the value of photosynthetically active radiation (PAR), its interception by leaves and quantum efficiency of the canopy, which can be modified by environmental conditions. The GPP is proportional to the radiation absorbed by the canopy, which is determined from the leaf area index (LAI) and PAR through Beer's law. The PAR that falls upon the canopy of the trees is calculated from the total incident solar radiation, being the quantum efficiency of the canopy reduced taking into account various environmental modifiers that are function of the vapour pressure saturation deficit, available water in the soil, air average temperature, number of frosty days per month, fertility and stand

age (Sands and Landsberg, 2002). The net primary productivity (NPP or PN) is a constant fraction γ (= 0.47) of gross primary productivity (Waring *et al.*, 1998).

2. Biomass/carbon allocation

The NPP allocation to the roots is determined by growth conditions that are expressed on the availability of water in the soil, vapour pressure deficit and site fertility, according to well-established principles (Beets, 1996; Landsberg and Gower, 1997). The fraction of NPP allocated to roots increases as soil fertility and/or the availability of water in the soil decrease. The biomass allocation to leaves and trunk with branches varies with growth conditions and depends on the average size of the trees, so that the biomass allocation to leaves decreases and allocation to the trunk and branches increases with the age of the population. Landsberg and Waring (1997) originally determined the value of this allocation ratios considering the average diameter and using the allometric functions for the average biomass of leaves, branches and stem. The ratios of biomass allocation are given considering the minimum and maximum ratio of allocation to roots, the ratio between the allocations to the leaves and woody materials and the effect of fertility on allocation - through the FR parameter.

3. Evolution of the number of trees / mortality

Population mortality may be density independent (random or stress-induced) or density dependent (self-thinning). An age dependent probability of death can be applied monthly and is potentially modified by factors in the long term. Variations in density are also calculated using the self-thinning law of 3/2 to ensure that the average stem biomass does not exceed the maximum allowed for each density value (Sands, 2003). For every tree that dies, a fraction of the average biomass of the total biomass reserve is removed. Since mortality alters both the number of trees and the biomass of each component, it may be necessary to use an iterative application of the self-thinning law to ensure that the law is fulfilled for the new state. When there is tree removal, it is assumed that each trunk removed has 20% of the average stem biomass and that no leaf loss occurs. This simulates the fact that the trees that die are usually the weakest and with very few small leaves (Sands and Landsberg, 2002).

4. Soil water balance

The 3-PG includes a simple soil water balance model that runs on a monthly cycle. To monthly rainfall (plus irrigation) opposes evapo-transpiration, which is calculated by the Penman-Monteith equation (Landsberg and Gower, 1997). The canopy interception is a fixed percentage of rainfall, but it can, from a certain value on, be a function of leaf area index. Excess soil water is lost as runoff. It is assumed that the vapour pressure deficit, available soil water and the age of the stand affect stomatal conductance. The canopy conductance is

determined from a potential stomatal conductance which is adjusted by the age modifier and the lowest of the environmental modifiers and increases with the increase in leaf area index (LAI) to the maximum conductance of the canopy (gCx (s m-1)).

5. Conversion of biomass into variables of interest for management

Starting from biomass values and assumptions about LAI, allocations and wood density, we obtain estimates of density, basal area, volume and other variables. NPP is partitioned in to the different biomass pools (foliage, aboveground woody tissue and roots) according to partition rates that depend on site and growth conditions and stand diameter at breast height. Litter-fall, root-turnover and wood mortality are also taken into account, but dynamic changes in allocation that are observed in thinning or pruning responses are not reproduced because allocation is dependent on tree size. Stem volume is calculated from stem biomass and using wood density while basal area is predicted considering the woody biomass and stand density. The outputs can be monthly or yearly and include biomass values of wood, roots and leaves, soil water, transpiration of the stand, leaf area index, volume, breast height diameter, average annual growth in volume and number of trees per hectare.

The initialization data such as latitude, density or even climatic data are, in principle, easy to obtain. The maximum available water in the soil (ASW) is difficult to obtain, because a correct assessment implies the opening of a profile pit, but if this requirement is met, it is easy to obtain. The fertility rating may pose a challenge, since its value has a high degree of uncertainty (in some applications this parameter has been manually adjusted according to available data, but this procedure cannot be implemented in an operational way or when no data are available). It is difficult to characterize the fertility of a site and data about soil profiles is not always available. However, the fertility and ASW both affect the light-use efficiency of the canopy, canopy conductance and the balance between above and below ground biomass allocation. It is therefore important to have good evaluation of these site variables.

The fertility ratio (FR) is a simple way to take into account the soil fertility. It varies from 0 for the most nutrient limited sites up to 1 for the sites at which nutrients are not limiting growth. The effects of site fertility on net primary production are taken into account by assuming that light use efficiency declines with declining FR, while biomass allocation to roots is assumed to increase with declining FR. Although these simple assumptions lead to realistic simulations of stand growth under a wide range of conditions, no simple objective means for assigning FR on the basis of soil chemistry has been available. So, this is a parameter that should be adjusted based on the soil knowledge, because not only is important in allocation, but Landsberg and Waring (1997) emphasized that although there is evidence to suggest that nutrition affects the amount of carbon allocated to roots, the quantification of this relationship

is difficult. There are statements that suggest that the nutritional status has a significant effect on productivity, having been suggested that the introduction of a sub-soil model could overcome these problems (e.g. Xenakis *et al.*, 2008) Furthermore, intensive forestry has shown that an increase in productivity can be achieved improving soil nutrition, because an increase in fertility increases the leaf area index (LAI).

The assignment of a value of FR to use in the 3-PG model has been widely discussed. Usually a value is assigned in a subjective way considering the soil chemistry and the knowledge of growth in the site or in similar sites. If there is a set of parameters for the species in sites that are not nutrient limited that are reliable, 3-PG can be used to determine the FR for each site by changing its value in the model and finding the one that better approximates the estimated and observed values of stand growth. These values can then be related to soil characteristics and try to see if there are any relations. FR is supposed to characterize site´s fertility, so it must take the same value in a specific site for different species. If different values arise, then there must be some species-specific differences that are not being considered in other model parameters.

There are studies that show the strengths and weaknesses of 3-PG (Sands and Landsberg, 2002; Landsberg *et al.*, 2003, 2005; Stape *et al.*, 2004) and even exploring the internal feedbacks between the parameters and outputs (Almeida *et al.*, 2004; Esprey *et al.*, 2004). Some aspects of the model are criticized, such as the FR and the self-thinning subroutine (Dye, 2001; Landsberg *et al.*, 2001; Whitehead *et al.*, 2002; Landsberg *et al.*, 2003). Landsberg *et al.* (2003) suggested the use of soil studies to estimate FR, but this method is not practical and has many limitations. In the same study it was also suggested the introduction of a dynamic model of soil organic matter to overcome this problem and this suggestion has been followed in several subsequent studies with the integration of 3-PG with models of soil organic matter decomposition (Peng *et al.*, 2002; Hirsch *et al.*, 2004; Paul and Polglase, 2004), although none has achieved an integration that allows the FR to be estimated from representations of soil processes. Although there is a rather limited knowledge of the relationship between the FR and the soil properties (Landsberg *et al.* 2003), efforts have been made in linking nutrient cycles with 3-PG, as well as advances in the development of local empirical functions between FR and soil characteristics. Stape *et al.* (2004) made a critical analysis using paired plots in eucalyptus plantations across 40 sites that captured the regional soil and productivity differences. First, 3-PG was parameterized using a complete carbon balance from an irrigated plantation and then validated using growth data from a set of 40 independent paired fertilized and unfertilized plots. The paired plots where of the same age and had similar starting values of woody biomass but one of them was fertilized to eliminate any existing limitation on tree growth. There were two growing seasons considered in the

study, one was normal and the other one was very wet. A fertilizer response (FER, t ha^{-1}) that measures the response of a site to fertilization was determined from the observed growth data for the wet years. 3-PG was also applied to the paired sites and the value of FR adjusted so that the predicted wood increment matched the observed increment in the wet years. When these values were compared with the observed FER, it was found that FR≈1 for sites with no fertilizer response (FER≈0) and that FR declined with increasing FER. The relationship between FER and soil parameters was also observed and a soil fertilizer response index (SFRI) was built, considering a linear combination of extractable potassium, phosphorus and cation exchange capacity in the upper soil layer that explained 56% of the observed variation in FER, but it is not known if the SFRI relationship is general. The study demonstrated that FR and the manner in which site fertility effects are included in 3-PG, enable it to predict observed responses to site fertility. Almeida *et al.* (2010) presented a model of FR as a function of soil fertility and available soil water, also considering other factors such as slope and management effects. In the work of Vega-Nieva *et al.* (2013), several plots were selected on which the soil characteristics were studied, including a full description of the soil profile, texture, pH, organic matter and fertility analysis. For each plot, the value of the available soil water was calculated with the model proposed by Domingo Santos *et al.* (2006). And the FR that best reproduced the observed growth was selected. A model for predicting the value of FR from the amount of available soil water and other nutrients with significant effects on fertility, in this case phosphorus, potassium, calcium and magnesium, was developed.

Forestry is primarily focused in selecting and applying the best set of silvicultural treatments needed to meet the objectives. So, one of the key uses of forest growth and yield models is projecting the long-term consequences of the various silvicultural decisions. Stand responses to treatment produce both direct and indirect effect (Miller and Tarrant, 1983; Auchmoody, 1985). In the context of fertilization, Opalach and Heath (1988) defined direct effect as "... the part of the response due to improved nutrition..." and the indirect effect as "... the remaining portion of the response due to altered stocking brought on by fertilizer in previous growing seasons." Ignoring the difference between direct and indirect effects of a treatment can lead to unreliable estimates of the size and duration of the treatments response (Auchmoody, 1985). Likelihood of a fertilization response does vary between stands (Peterson *et al.*, 1986) and it can be difficult to predict. The size of the direct effect response depends on tree species, type and amount of fertilizer, time since applications and the site productivity, with poor-quality sites often having larger relative responses than the most high-quality sites.

Plants require a number of mineral nutrients for their growth and inadequate supplies will affect it. Nitrogen is ultimately derived from atmospheric nitrogen, but all the other nutrients have their origin in the soil material. The macro-nutrients: nitrogen (N), phosphorus (P), sulphur (S),

calcium (Ca), magnesium (Mg), and potassium (K) are required in relatively large quantities. The micro-nutrients or trace elements are boron (B), chlorine (Cl), cobalt (Co), copper (Cu), iron (Fe), manganese (Mn), molybdenum (Mo), nickel (Ni), and zinc (Zn).

The concept of site fertility is complex, it is the outcome of many factors rooted in soil chemistry; it varies both seasonally and throughout the lifetime of a forest stand as a consequence of changing climatic conditions and it is affected by the withdraw of nutrients following litter fall. Models that could fully account for all these various factors would be extremely complex. However, some attempts have been made to avoid these complexities by devising pragmatic fertility indices, such as the 3-PG fertility rating.

Forest site quality is defined as the physical and biological factors that characterize a site´s ability to support tree growth (Skovsgaard and Vanclay, 2008). One of the most common measures of site quality is Site index (S), which is the expected height of the trees at a certain reference age. S has a long story of development and use, as its utility was noted towards the end of the nineteenth century (Batho and García, 2006). Some of the reasons for its widespread use are the relative ease to be estimated from field observations, its proven efficiency in predicting volume growth and yield and the strength of the site-specific relationship between dominant height and age in even-aged forests. Some of the key assumptions behind the use of S as a measure of site quality are:

- dominant height is independent of stand conditions;

- dominant trees that have not experienced any suppression or other damage are easily identified:

- dominant height is an effective integrator of the key biological determinants of growth.

Although these assumptions are generally met, there are some significant exceptions, which can limit the utility of S as a productivity measure. For example, several studies found that tree height to be significantly influenced by stand density (Lynch, 1958; Cieszewsky and Bella, 1993; MacFarlane *et al.*, 2000; Flewelling *et al.*, 2001). The severity of the influence is related to species tolerance and site conditions; however, adjustments to ensure unbiased estimates can be made (e.g. Flewelling *et al.*, 2001). To make sure that S is representative of the site quality, only undamaged trees must be selected, but those are usually the dominant trees.

S is based on a reference age, which is typically set at some age less than the anticipated rotation age (Goelz and Burk, 1996). S estimates accuracy decreases significantly as stands diverge from the S reference age and Goelz and Burk (1992) found that top height at S age is often underestimated for high quality sites and overestimated for low quality sites. Determination of S requires the measurement of height and age of individual trees, both of

which are subjected to error. Measurement errors bias the S estimates and it can also bias the S equations (Goelz and Burk, 1996).

The ideal measure of site productivity is consistent over time, and for a given stand, S varies due to changes in genetics, climate and management practices. Monserud and Rehfeldt (1990) found that genetic variation explained over 40% of the original variation in S, which suggests that S can change between generations. Environmental conditions also have a significant influence on S estimates, because they affect growth on the short-term and stand genetics in the long-term (Monserud and Rehfeldt, 1990). Some silvicultural treatments like soil bedding can increase site productivity permanently, but other like vegetation control or fertilization just accelerates stand development without changing the final volume (Fox, 2000). When stands are thinned, measures of site productivity based on stand height have reduced ability (Skovsgaard and Vanclay, 2008). Even when the practices have long-term consequences, S is very sensitive to forest management practices. The drawbacks of the S as a site productivity measure give impetus to the use of process-based models that directly use environmental variables to drive tree and stand growth.

I.3. Objectives and outline

The work described here is a continuation of the work that has been being developed by the author in the last years. This work started with the GlobLand project. The work in that project gave the insight not only of the GLOBULUS model, the equations developed and all the knowledge applied in them, but it also gave a good insight of the eucalyptus stands data available. The work developed in the author's MSc thesis took advantage of that knowledge and a system of equations for total and components above ground biomass of eucalyptus plantations was developed. This work also revealed some bias in the existing crown ratio (cr) equation for young stands, which was the starting point of the work described here.

The purpose of this work is to improve the forest models available to support eucalyptus management – GLOBULUS and 3-PG – and to combine/hybridize them in order to obtain a model to support eucalyptus plantations management under a changing environment. This new model must be of simple use and with multiple outputs; sensible to climate changes; be capable of considering different management alternatives and account for different management options such as fertilizations. After an initial stage of analysis of the existing models, the following partial objectives where defined:

1. Improve biomass estimations (the basis of process-based models development) by:
    a) Readjusting the crown ratio equations

b) Develop a stand level system of equations to estimate total aboveground biomass and biomass per tree components (wood, bark, branches and leaves)

2. Improve the output of the 3-PG model so that it provides the same information as GLOBULUS – hybridization

3. Model diameter distributions

4. Develop a prediction model for the fertility index FR from 3-PG using soil characteristics as regressors

Models have to be accurate to be useful and accuracy can be improved by new data collection, improvement of by the quality of the existing data or the use of more sophisticated modelling techniques to the existing data. The need to readjust the crown ratio equations arose when the available data for eucalyptus stands in Portugal was analysed by Oliveira (2008). This author found that, in smaller trees, the difference between the measured and estimated values for crown length was very significant and the improvement of the equation was suggested at that time.

Any improvement to be made in the 3-PG model has to be based on a precise database with information of several variables, namely biomass values. Being the crown ratio a fundamental variable for the biomass estimates, the starting point of this work was the improvement of those equations. The previous equations (Soares and Tomé, 2001) were fitted using the limited data that was available at the time, and they covered only a few age classes and localized areas of Portugal. The eucalyptus database that is currently available has a much higher number of data covering a wide range of ages, locations, and other features of the stand, allowing the development of a better equation. The availability of measurements made in consecutive years allowed the development of an equation for crown length change and also to consider the autocorrelation structure of the data, something that had not been done previously. To overcome compatibility issues, a simultaneous adjustment was also made with a crown length equation. The new equations improved the biomass estimations in the database, especially in terms of leaves and branches biomass, and lead to the need to improve the existing stand level biomass equations. Once the biomass equations were improved, the development of a model to support forest management of eucalyptus plantations in a changing environment, the main objective of this thesis, could take place.

To satisfy the need to respond to climate changes, the growth models used have to be process-based (Lansberg, 2003; Battaglia *et al.*, 2007; Pretzsch *et al.*, 2008). The reasons why the 3-PG (Physiological Principles in Predicting Growth) model (Landsberg and Waring, 1977, Sands and Landsberg, 2002) was chosen where: 3-PG is a process-based model, it had

already been calibrated for eucalyptus stands in Portugal (Fontes *et al.*, 2006) and it is a quite simple and not very demanding model in what concerns site characterization for model initialization, which is an advantage to be considered. In Portugal it had already been complemented with some prediction functions that improve the quality of the model output (Tomé *et al.*, 2004).

3-PG is stand level model that works on a monthly basis (although it has monthly or annual outputs) and that includes empirical and process-based relationships that make the link between growth models based in measurements and process-based carbon balance models. It calls for stand, soil and climate data as initial inputs and supplies dynamic predictions of biomass, stand basal area (G) and density (N), among other variables. The main advantages, regarding other available models, namely empiric ones, is the ability to be used in climate changes situations and to be adapted to different species through parameterization with existing stand level data. Nevertheless, it is not as easy to understand as empirical growth models, not necessarily precise, it needs data that might not be available, and it gives a less detailed level of output and it makes a coarse approach to soil fertility.

The objectives of the hybridization are the improvement of the predictions of stand basal area (G), dominant height (hdom) and biomass and the diversification of outputs. This process is an attempt to combine different types of models, one process-based that can reflect management and climate changes and an empirical model that can give detailed outputs.

The 3-PG "information for managers" module has 3 sub-models: stand density, basal area and volume under bark. Stand density is predicted according to the 3/2 power law while basal area and volume under bark are predicted based on woody biomass estimations.
The calculation of the volume under bark is based on the prediction of woody biomass, one of the outputs of 3-PG. The woody biomass (Wwoody = wood biomass + branches biomass + bark biomass) is converted into wood biomass (Ww) through the ratio (R_{rc}) between branches and bark and woody biomass which is then multiplied by the wood density (ρ) to give an estimate of the volume under bark (V_{u_st}):

$$V_{u\_st} = W_{woody}\ R_{rc}\ \rho$$

Both R_{RC} and ρ are modelled as a function of age.

Basal area estimates takes a few more calculation steps and has several associated problems. First, the woody biomass of the mean tree is predicted dividing the total woody biomass by the stand density; then, the allometric equation for tree woody biomass is “inverted” and used to estimate the quadratic mean diameter (dg) that is used to compute the basal area of the mean

tree (g_{med}). Basal area is then estimated multiplying the basal area of the mean tree by the stand density:

$$\overline{W_{woody}} = \frac{W_{woody}}{N}$$

$$\overline{W_{woody}} = k_{Ww}\, dg^{a\,Ww} \Rightarrow\ dg = \left(\frac{\overline{W_{woody}}}{k_{Ww}}\right)^{1/a\,Ww} \Rightarrow\ g_{med} = \frac{\pi}{4}\left(\frac{dg}{100}\right)^2$$

$$G = N\, g_{med}$$

The aim is to improve this module, diversifying the outputs provided by 3-PG by hybridization with the empirical model GLOBULUS to obtain an output with the same level of detail. The connection between the two models is achieved by the allometric relationship that exists between basal area and woody biomass (Tomé *et al.*, 2004). The biomass growth in the period under consideration is calculated by 3-PG and the present value of woody biomass ($Wwoody_{t2}$) is calculated by adding the value of growth to the previous value. The prediction value of basal area (G_{t2}) is then estimated based on the existing allometric relationship between basal area and woody biomass expressed as a difference equation (G-link function):

$$G_{t2} = k\, Wwoody_{t2}{}^{a} \Rightarrow G_{t2} = G_{t1}\left(\frac{Wwoody_{t2}}{Wwoody_{t1}}\right)^{a}$$

with k - constant, a – allometric parameter and t1 and t2 the beginning and end of the growing period.

The dominant height, which is usually used as a regressor for the estimation of volume in growth models, is estimated by an allometric relationship with the total aboveground biomass (Wa) in which the allometric parameter is a function of density (hdom-link):

$$hdom_{t2} = k\, W_{a2}{}^{a0+a\,N\,t2\,/1000} \Rightarrow hdom_{t2} = hdom_{t1}\left(\frac{W_{a2}{}^{a0+a\,N\,t2\,/1000}}{W_{a1}{}^{a0+a\,N\,t1\,/1000}}\right)$$

With the estimates of basal area and dominant height all the GLOBULUS output can be obtained. Diameter distribution is very important in the stand description. All other variables such has height, volume and biomass are correlated with the tree diameter. The diameter distribution simulation is also essential for the initialization of individual tree models (Ek and Monserud, 1979) especially when there is no data available or when there is the need to simulate the transition from a regular to an irregular stand, which implies the use of individual tree models. Besides, it is also essential in the planning of efficient harvesting operations. For the simulation of the diameter distribution, prediction equations for the minimum, average,

median and maximum diameters were developed based on stand level variables such has dominant height, density, site index and quadratic mean diameter. These variables allow the use of the available algorithm that simulates the diameter distribution (Marto *et al.*, 2009) with Johnson's probability density function selected by Mateus and Tomé (2011).

The last objective of this thesis is the improvement of one of the weaknesses already pointed to the 3-PG model, the estimation of the fertility ratio (FR). The FR is one of the inputs of the model and it is a value that varies from 0 to 1 that is assigned in a somehow empirical way according to the site soil characteristics. The idea was to develop a model that predicts the FR value considering soil characteristics and fertility, which would allow the model to respond to fertilization changes. To make this improvement, a model that relates the FR with soil characteristics was developed. This will also allow to find out what are the most limiting nutrients for eucalyptus productivity in Portugal. The work developed by Vega-Nieva *et al.* (2013) showed good results for Galicia and similar results were expected for Portugal. However, the development of a model to predict FR in Portuguese eucalyptus plantations was not as easy as it seemed to be in Galicia.

I.4. Data

There were 2 different sets of data used in this thesis, one set from the large database available in Portugal for eucalypt plantations and the other one from 5 different fertilization trials belonging to the pulp company ALTRI.

The first set, the large database available in Portugal for eucalypt plantations includes data from permanent plots, continuous forest inventory of the pulp companies, spacing trials, and data from an experiment to study the species potential productivity that includes fertilization and irrigation as treatments. The plots are located throughout the species area of distribution in Portugal (Figure 1) and represent not only the existing ranges of ages, stand densities, sites and management options, but also give information of less usual management options, such as high densities and older trees, since the usual rotation cycle for eucalyptus doesn´t tend to go beyond 11/12 years.

It contains data from first-rotations and coppiced stands, part of it consisting of successive measurements (usually with 1 year interval between them) on the same plots which means that the presence of auto-correlation has to be considered. On each plot every tree had its diameter at breast height (d) measured. In current inventories, usually only trees with d bigger than 5cm are measured, but on permanent plots and trials all trees are measured, which means that there is information on very small trees.

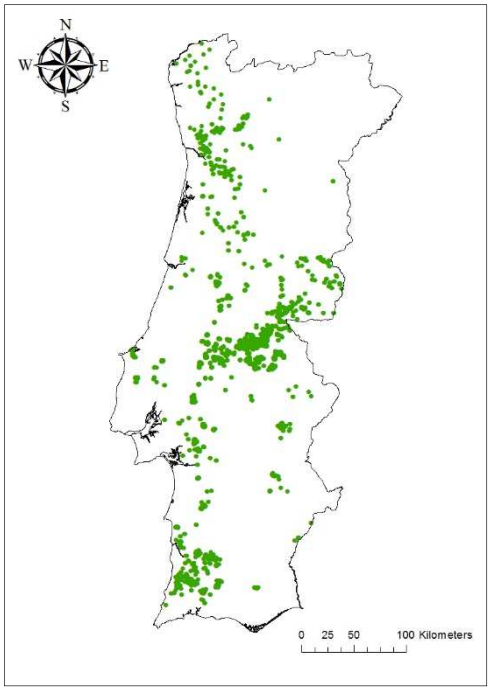


Figure 1. Plots distribution.

The plot size ranged from 100 to 2464 m2, depending on the source of the data. Continuous forest inventory plot size is usually 400-500 m2, but in permanent plots and trials, plot size varies. Diameter at breast height was measured in every tree and total tree height in dominant, model trees or all the trees, depending on data provenience. Non-existing tree heights were estimated with adequate allometric equations. Different stand volumes – total volume without stump over and under bark, bark volume, merchantable volumes to top diameters between 10 and 5 cm – were computed using the system of volume equations from Tomé *et al.* (2007). Aboveground biomass and biomass per tree component (stem wood, stem bark, branches and leaves) were estimated using the system of biomass equations developed by António *et al.* (2007). Crown length, used as regressor in the leaf and branches biomass equations, was estimated with the crown ration and crown length system of equations developed by Oliveira and Tomé (submitted). Site index was estimated with the equation from Tomé *et al.* (2001).

Stand variables for each plot included age (t, years), stand density (N, trees ha-1), basal area (G, m2 ha-1), dominant height (hdom, m), site index (S, m) and total aboveground biomass (Wa) and biomass per tree component (Wi, Mg ha-1, i=w for stem wood, b for stem bark, br for branches and l for leaves).

The second set of data came from 5 different properties belonging to the pulp company ALTRI, scattered along the eucalyptus distribution area in Portugal (Figure 2) and in which fertilization trials are installed. All plots are of first rotation stands of non-clonal *Eucalyptus globulus*,

planted at 1.8mx4m spacing and were measured between the ages of 2 to 8.8 years. Each plot had an area of 518,4 m^2 with a double edge and a number of 12 or 16 useful plants for height and diameter measurements.

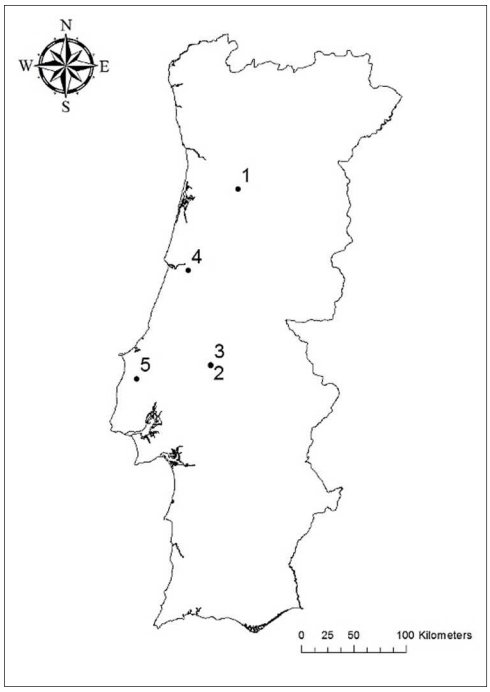


Figure 2. Trials distribution.

References

AFN - Autoridade Florestal Nacional. 2010. Inventário Florestal Nacional. IFN 2005-2006. Portugal Continental. Autoridade Florestal Nacional, Ministério da Agricultura do Desenvolvimento Rural e das Pescas, Lisbon, Portugal.

Ågren, G.I. 1996. Nitrogen productivity of photosynthesis minus respiration to calculate plant growth. Oikos 76:529-535

Almeida, A.C., Landsberg, J.J., Sands, P.J. 2004. Parameterisation of 3-PG model for fast-growing Eucalyptus grandis plantations. Forest Ecology and Management. 193, 179–195.

Almeida, A.C., Maestri, R., Landsberg, J.J., Scolforo, J.R.S. 2003. Linking process-based and empirical forest models in Eucalyptus plantations in Brazil. In "Modelling Forest Systems", p. 63-74, Editors A. Amaro, D. Reed and P. Soares, CAB International.

Almeida, A.C., Siggins, A., Batista, T.R., Beadle, C., Fonseca, S., Loos, R. 2010. Mapping the effect of spatial and temporal variation in climate and soils on Eucalyptus plantation production with 3-PG, a process-based growth model, Forest Ecology and Management. 259(9): 1730-1740.

Amaro, A., 2003. SOP model. The SOP Model: the Parameter Estimation Alternatives. In: Amaro, A., Reed, D. and Soares, P.(Eds.), Modelling Forest Systems. CABI Publishing, USA.

Amaro, A., Reed, D., Tomé, M., Themido, I. 1998. Modelling Dominant Height Growth: Eucalyptus Plantations in Portugal. Forest Science, 44 (1): 37-46.

Amatya, D.M., Skaggs, R.W. 2001. Hydrologic modelling of a drained pine plantation on poorly drained soils. Forest Science. 47(1): 103–114

António, N., Tomé, M., Tomé, J., Soares, P., Fontes, L. 2007. Effect of tree, stand and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research 37: 895-906.

Auchmoody, L.R. 1985 Evaluating growth responses to fertilization. Canadian Journal of Forest Research 15: 877-880.

Baldwin, V.C.Jr., Burkhart, H.E., Westfall, J.A., Peterson, K.D. 2001. Linking growth and yield and process models to estimate impact of environmental changes on growth of loblolly pine. Forest Science. 47: 77-82.

Barreiro, S., Tomé, M., Tomé, J. 2004. Modelling growth of unknown age even-aged eucalyptus stands. In: Hasenauer, H., Mäkelä, A. (Eds.), Modelling Forest Production. Scientific Tools—Data Needs and Sources. Validation and Application. Proceedings of the International Conference, Wien, pp. 34–43.

Bartelink, H.H. 2000. Effects of stand composition and thinning in mixed-species forests: a modelling approach applied to Douglas-fir and beech. Tree Physiology. 20: 399–406.

Batho, A., García, O. 2006. De Perthius and the origin of site index: a historical note. Forest Biometry, Modelling and Information Sciences 1: 1-10.

Battaglia M., Sands, P.J. 1998. Process-based forest productivity models and their application in forest management. Forest Ecology and Management 102: 13-32.

Battaglia, M., Almeida, A., O'Grady, A. & Mendham, D. 2007. Process-based models in Eucalyptus plantations management: reality and perspectives. Boletín del CIDEU 3: 189-205.

Battaglia, M., Sands, P.J., Candy, S.G. 1999. Hybrid growth model to predict height and volume growth in young *Eucalyptus globulus* plantations. Forest Ecology and Management. 120: 193-201.

Beets, P.N.W., D. 1996. Carbon partitioning in *Pinus radiata* stands in relation to foliage nitrogen status, Tree Physiol. 16: 131-138.

Botkin, D. B., Janak, J. F., and Wallis, J. R. 1972. Some Ecological Consequences of a Computer Model of Forest Growth. Journal of Ecology. 60: 849–872.

Botkin, D.B. 1993. JABOWA-II: a computer model of forest growth, Oxford University Press, New York, NY.

Brunner, A., Nigh, G. 2000. Light absorption and bole volume growth of individual Douglas-fir trees. Tree Physiology. 20: 323–332.

Canham, C.D., Coates, K.D., Bartemucci, P., Quaglia, S. 1999. Measurement and modelling of spatially explicit variation in light transmission by canopy trees. Canadian Journal of Forest research, 29: 1775-1783.

CELPA - Associação da Indústria Papeleira. 2011. Boletim Estatístico 2010 da Associação da Indústria Papeleira. CELPA Editores. Lisbon, Portugal.

Cieszewsky, C.J., Bella, I.E. 1993 Modelling density-related lodgepole pine height growth using Czarnowsky´s stand dynamic theory. Canadian Journal of Forest research 23: 2499-2506.

Coops, N.C., Waring, R.H., Landsberg, J.J. 1998. Assessing forest productivity in Australia and New Zealand using a physiologically-based model driven with averaged monthly weather data and satellite derived estimates of canopy photosynthetic capacity. Forest Ecology and Management, 104, 113-127.

Courbaud, B. 2000. Comparing light interception with stand basal area for predicting tree growth. Tree Physiology. 20: 407–414.

DGSFA - Direcção Geral dos Serviços Florestais e Aquícolas, 1966a. Inventário ao Norte do Tejo - 1965-66. Direcção Geral dos Serviços Florestais e Aquícolas, Lisboa, Portugal.

DGSFA - Direcção Geral dos Serviços Florestais e Aquícolas, 1966b. Inventário ao Sul do Tejo - 1965-66. Direcção Geral dos Serviços Florestais e Aquícolas, Lisboa, Portugal.

Ditzer, T., Glauner, R., Förster, M., Köhler P., Huth A. 2000. The process-based stand growth model FORMIX 3-Q applied in a GIS environment for growth and yield analysis in a tropical rain forest. Tree Physiology. 20: 367–381.

Domingo Santos, J.M., Fernández de Villarán, R., Corral, E., Rapp, I. 2006. Estimación de la capacidad de retención de agua en el suelo: revisión del parámetro CRA. Invest Agraria: Sistemas y Recursos Forestales 15:14–23.

Dye, P.J., 2001. Modelling growth and water use in four *Pinus patula* stands with the 3-PG model. South African Forestry Journal. 19: 53–63.

Ek, A.R., Monserud, R.A. 1979. Performance and comparison of stand growth models based on individual tree and diameter class growth. Canadian Journal of Forest Research. 9: 231–244

Esprey, L.J., Sands, P.J., Smith, C.W. 2004. Understanding 3-PG using a sensitivity analysis. Forest Ecology and Management. 193: 235–250.

Flewelling, J., Collier, R., Gonyea, B., Marshall, D., Turnblom, E. 2001. Height-age curves for planted stands of Douglas-fir, with adjustements for density. Stand Management Cooperative Working Paper 1. University of Washington, College of Forest Resources, Seattle, WA.

Fontes, L., Landsberg, J., Tomé, J. Tomé, M., Pacheco, C. A., Soares, P., Araújo, C. 2006. Calibration and testing of a generalized process-based model for use in Portuguese eucalyptus plantations. Canadian Journal of Forest Research. 36: 3209-3221.

Ford, A. 1999. Modelling the Environment, Island Press, Washington, DC.

Fox, T.R. 2000. Sustained productivity in intensely managed forest plantations. Forest Ecology and Management 138: 187-202

Garcia-Gonzalo, J., Peltola, H., Briceño-Elizondo, E., and Kellomäki, S. 2007. Changed thinning regimes may increase carbon stock under climate change: a case study from a Finnish boreal forest. Climatic Change, 81(3–4): 431–454.

Goelz, J.C.G., Burk, T.E. 1996. Measurement error causes bias in site index equations. Canadian Journal of Forest Research 26: 1585-1593.

Green, E.J., MacFarlane, D.W., Valentine, H.T., Strawderman, W.E. 1999. Assessing uncertainty in a stand growth model by Bayesian synthesis. Forest Sciences, 45: 528-538.

Hirsch, A.I., Little, W.S., Houghton, R.A., Scott, N.A., White, J.D. 2004. The net carbon flux due to deforestation and forest re-growth in the Brazilian Amazon: analysis using a process-based model. Global Change Biology. 10: 908–924.

Isebrands, J.G, Rauscher, H.M, Crow, T.R., Dickman, D.J. 1990. Whole tree growth process models based on structural-functional relationships. In: Dixon, R.K., Meldhal, R.S., Ruark, G.A., Warren, W.G. (Eds), Forest Growth: Process Modelling of Forest GrowthR to Environmental Stress. Timber Press, Portland, OR. pp 96-112.

Jakeman, A.J., Chen, S.H., Rizzoli, A.E., Voinov, A.A. 2008. Modelling and software as instruments for advancing sustainability, in In: Jakeman, A.J., Voinov, A.A., Rizzoli, A.E., Chen, S.H (eds). Environmental Modelling, Software and Decision Support: State of the Art and New Perspectives, Elsevier, Amsterdam.

Johnsen, K., Samuelson, L., Teskey, R., McNulty, S., Fox, T. 2001. Process models as tools in forestry research and management. Forest Science 47: 2-8.

Kimmins, J.P., Mailly, D., Seely, B. 1999. Modelling forest ecosystem net primary production: the hybrid simulation approach used in FORECAST, Ecological Modelling, 122: 195- 224.

Kirschbaum, M.U.F. 2000. Forest growth and species distribution in a changing climate. Tree Physiology. 20: 309–322.

Korzukhin, M.D., Ter-Mikaelian, M.T., Wagner, R.G. 1996. Process versus empirical models: which approach for forest management? Canadian Journal of Forest Research. 26: 879-887.

Landsberg, J.J. 1986. Physiological ecology of forest production. Academic Press. New York, NY, 198 p.

Landsberg, J.J. 2003. Physiology in forest models: history and the future. Forest Biometry Modelling and Information Sciences, volume 1: 49-63.

Landsberg, J.J. 2003b. Modelling forest ecosystems: state of the art, challenges and future directions. Canadian Journal of Forest Research. 33: 385-397.

Landsberg, J.J., Gower. S.T. 1997. Applications of Physiological Ecology to Forest Management. Academic Press, London, p. 354.

Landsberg, J.J., Johnsen, K.H., Albaugh, T.J., Allen, H.L., McKeand, S.E. 2001. Applying 3-PG, a simple process-based model designed to produce practical results, to data from Loblolly Pine experiments. Forest Science. 41: 43–51.

Landsberg, J.J., Kaufmann, M.R., Binkley, D. Isebrands, J., Jarvis, P.G. 1991. Evaluating progress toward closed forest models based on fluxes of carbon, water and nutrients. Tree Physiology. 9: 1–15

Landsberg, J.J., Mäkelä, A., Sievänen, R., Kukkola, M. 2005. Analysis of biomass accumulation and stem size distributions over long period in managed stands of *Pinus sylvestris* in Finland using the 3-PG model. Tree Physiology. 25:, 781–792.

Landsberg, J.J., Waring, R.H., Coops, N.C. 2003 Performance of the forest productivity model 3-PG applied to a wide range of forest types. Forest Ecology and Management. 172:, 199–214.

Landsberg, J.J.,Waring, R.H. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning. Forest Ecology and Management. 95: 209–228.

Law, B.E., Goldstein, A.H., Anthoni, P.M., Unsworth, M.H., Panek, J.A., Bauer, M.R., Fracheboud, J.M., Hultman, N. 2001. CO2 and water vapor exchange by young and old ponderosa pine ecosystems during a dry summer. Tree Physiology. 2: 299–308.

Leary, R.A. 1985 A framework for assessing and rewarding a scientist´s research productivity. Scientometrics 7: 29-38.

Levins, R. 1966. The strategy of model building in population biology. Am Scientists 54: 421-431.

Levy, P.E., Hale, S.E., and Nicoll, B.C. 2004. Biomass expansion factors and root:shoot ratios for coniferous tree species in Great Britain. Forestry, 77(5): 421–430.

Lindner, M. 2000. Developing adaptive forest management strategies to cope with change. Tree Physiology. 20:299–307.

Lynch, D.W. 1958. Effects of stocking on site measurement and yield of second-growth ponderosa pine in the Inland Empire. Research Paper 56. USDA Forest Service, Intermountain Forest and Range Experiment Station, Oregon, UT.

MacFarlane D.W., Green, E.J., Burkhart, H.E. 2000. Population density influences assessment and application of site index. Canadian Journal of Forest Research 30: 1472-1475.

Mäkelä, A, Landsberg, J.,Ek, A.R.,Burk, T.E.,Ter-Mikaellian, M.,Agren, G.I., Oliver, C.D., Puttonen, P. 2000. Process-based models for forest ecosytem management: Current state of the art and challenges for practical implementation. Tree Physiology. 20: 289–298.

Mäkelä, A. 1988. Performance analysis of a process-based stand growth model using Monte Carlo techniques. Scandinavian Journal of Forest Research. 3: 315-331.

Mäkelä, A. 2009. Hybrid Models of forest stand growth and production. In Dykstra, D.P. and Monserud, R.A. (eds) Forest Growth and Timber quality; Crown Models and Simulation Methods for Sustainable Forest Management, US Department of Agriculture, Forest Service, Pacific Nortwest Research Station, Portland, OR.

Mäkelä, A., Makinen, H. 2003. Generating 3D sawlogs with a process-based growth model. Forest Ecology and Management. 184: 337-354.

Mäkelä, A., Sievänen, R., Lindner, M., Lasch, P. 2000a. Application of volume growth and survival graphs in the evaluation of four process- based forest growth models. Tree Physiology. 20: 347–355.

Marto, M., Palma, J., Mateus, A., Tomé, M. 2009. Computer program for estimation of Johnson´s S_B parameters using a parameter recovery method. Publicações Científicas Forchange PC-X/2009. Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa.

Matala, J., Hynynen, J., Miina, J., Ojansuu, R., Peltola, H., Sievänen, R., Väisänen, H., Kellomäki, S. 2003. Comparison of a physiological model and a statistical model for prediction of growth and yield in boreal forests. Ecological Modelling, 161: 95-116.

Mateus, A., Tomé, M. 2011. Modelling the diameter distribution of eucalyptus plantations with Johnson's probability density function: parameters recovery from a compatible system of equations to predict stand variables. Annals of Forest Science. 68(2): 325-335. DOI: 10.1007/s13595-011-0037-7.

McMurtrie, R.E., Wang, Y.P. 1993. Mathematical models of the photosynthetic response of tree stands to rising CO2 concentrations and temperature. Plant Cell and Environment. 16: 1–13.

Miller, R.E., Tarrant, R.F. 1983. Long-term growth response of Douglas-fir to ammonium nitrate fertilizer. Forest Science 29: 127-137.

Milner, K.S., Cobel, D.W., McMahan, A.J., Smith, E.L. 2003. FVSBGC: a hybrid of the physiological model STAND-BGC and the forest vegetation simulator. Canadian Journal of Forest Research, 33: 466-479.

Mohren, G.M.J., Burkhart, H.E. 1994. Contrasts between biologically-based process models and management-oriented growth and yield models. Forest Ecology and Management. 69: 1–5.

Monserud, R.A., 2003. Evaluating forest models in a sustainable forest management context. Forest Biometry Modelling and Information Sciences, vol 1, 35-47.

Monserud, R.A., Rehfeldt, G.E. 1990. Genetic and environmental components of variation of site index in inland Douglas-fir. Forest Science. 36(1): 1-9.

Oliveira, T.S. 2008. Sistema para a predição de biomassa aérea total e por componentes em povoamentos puros regulares de *Eucalyptus globulus* Labill. Tese de Mestrado em Engenharia Florestal e dos Recursos naturais, Instituto Superior de Agronomia, Lisboa.

Oliveira, T.S., Tomé, M. 2015. Modelling the aboveground stand biomass of planted and coppiced *Eucalyptus globulus* in Portugal. Submitted.

Opalach, D., Heath, L.S. 1988. Evaluation of long-term fertilizer response. In Ek, A.R, Shifley, S.R., Burk, T.E. (Eds.), Forest growth modelling and prediction. General Technical Report NC-120. USDA Forest Service, North Central Forest Experimental Station, St.Paul, MN, pp. 555-561.

Paul, K.I., Polglase, P.J. 2004. Calibration of the Roth C model to turnover of soil carbon under eucalypts and pines. Australian Journal of Soil Research 42(8): 883-95.

Peng, C.H., Liu, J., Dang, Q.L., Apps, M., Jiang, H., 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. Ecological Modelling. 153: 109–130.

Peterson, C.E., Webster, S.R., Barker, P.R., Miller, R.E. 1986. Using nitrogen fertilizers in management of coast Douglas.fir:II. Future informational needs. In Oliver, C.D., Hanley, D.P., Johnson, J.A. (Eds.), Douglas-fir: Stand management for the future. Contribution Number 55. University of Washington, Institute of Forest Resources, Seattle, WA, pp: 304-309

Pretzsch, H., Grote, R., Reineking, B., Rotzer, Th., Seifert, St. 2008. Models for Forest Ecosystem Management: A European Perspective. Annals of Botany. 101(8):1065-1087, May 2008.

Raulier, F., P.Y. Bernier, Ung, C.H. 2000. Modelling the influence of temperature on monthly gross primary productivity of sugar maple stands. Tree Physiology. 20: 333–345.

Robinson, A. P., Ek, A. R. 2003. Description and validation of a hybrid model of forest growth and stand dynamics for the Great Lakes region. Ecological Modelling. 170(1: 73-104. DOI: 10.1016/S0304-3800(03)00304-1

Ryan, M.G., Hubbard, R.M., Pongracic, S., Raison, R.J., McMurtrie, R.E. 1996. Foliage, fine-root, woody-tissue and stand respiration in *Pinus radiata* in relation to nitrogen status. Tree Physiology 16: 333-343.

Sands, P. J., Landsberg, J. J. 2002. Parameterisation of 3-PG for plantation grown *Eucalyptus globulus*. Forest Ecology and Management, 163: 273-292.

Sands, P., M. Battaglia., Mummery, D. 2000. Application of process-based models to forest management: experience with PROMOD, a simple plantation productivity model. Tree Physiology. 20: 383–392.

Santos, P.M., Soares, P., Mendes, A.M.S.C., Caldeira, B., Praxedes, J., Brígido, S., Pina, J.P., Paulo, J.A., Tomé, M., Barreiro, S., Palma, J.H.N., Borges, J.G.C., Garcia-Gonzalo, J., Sottomayor, M. 2013. Estudo prospetivo para o setor florestal - relatório final. AIFF - Associação para a Competitividade da Indústria da Fileira Floresta, 295 pp.

Schmid, S., Thurig, E., Kaufmann, E., Lischke, H., Bugmann, H. 2006. Effect of forest management on future carbon pools and fluxes: A model comparison. Forest Ecology and Management 237: 65-82.

Sharpe, P.J.H. 1990. Forest modelling approaches: compromises between generality and precision. *In* Process Modelling of Forest Growth Responses to Environmental Stress. Eds. R.K. Dixon, R.S.

Sievanen, R., Burk, T.E. 1992. Fitting process-based models with stand growth data: Problems and experiences. Forest Ecology and Management. 69: 145–156.

Skovsgaard, J.P., Vanclay, J.K. 2008. Forest site productivity: A review of the evolution of dendrometric conceptsfor even-aged stands. Forestry, 81: 13-31.

Smith, J.E., Woodbury, P.B., Weinstein, D.A. Laurance, J.A. 1998 Integrating research on climate change effects on loblolly pine: A probabilistic regional modelling approach. *In* The productivity and sustainability of southern forest ecosystems in a changing environment, Mickler, R.A., and S. Fox (eds.). Ecological Studies. 128: 429–451.

Soares P., Tomé, M. 2003. GLOBTREE: an Individual Tree Growth Model for Eucalpytus globulus in Portugal. In: Amaro, A., Reed, D., Soares, P. (Eds.), Modelling forest systems. CAB International, pp. 97-110.

Soares, P.,Tome, M. 2001. A tree crown ratio prediction equation for eucalypt plantations. Annals of Forest Science. 58(2): 193-202.

Stape, J.L., Ryan, M.G., Binkley, D. 2004. Testing the utility of the 3-PG model for growth of Eucalyptus grandis × urophylla with natural and manipulated supplies of water and nutrients. Forest Ecology and Management. 19: 219–234.

Taylor, A.R., Chen, H.Y.H., VanDamme, L. 2009. A review of forest succession models and theirs suitability for forest management planning. Forest Sciences, 55:, 23-36.

Thornley, J.H.M., Cannell, M.G.R. 1996. Temperate forest responses to carbon dioxide, temperature, and nitrogen: a model analysis. Plant Cell and Environment. 19: 1331–1348.

Tomé, M., Borges, J.G., Falcão, A. 2001. The use of Management-Oriented Growth and Yield Models to Assess and Model Forest Wood Sustainability. A case study for Eucalyptus Plantations in Portugal. In: Carnus, J.M., Denwar, R., Loustau, D., Tomé, M., Orazio, C. (Eds.), Models for Sustainable Management of Temperate Plantation Forests, European Forest Institute, Joensuu, pp. 81-94.

Tomé, M., Faias, S. P., Tomé, J., Cortiçada, A., Soares, P. & Araújo, C. 2004. Hybridizing a stand level process-based model with growth and yield models for *Eucalyptus globulus* plantations in Portugal. In: Borralho, N. M. G., Pereira, J. S., Marques, C., Coutinho, J.,

Madeira, M. & Tomé, M. (ed.), Eucalyptus in a changing world. Proc. Iufro Conf., Aveiro, 11-15 Oct. (RAIZ, Instituto de Investigação da Floresta e do Papel, Portugal), pp.290-297.

Tomé, M., Oliveira, T., Soares, P. 2006 O modelo GLOBULUS 3.0. Publicações GIMREF - RC2/2006. Universidade Técnica de Lisboa. Instituto Superior de Agronomia. Centro de Estudos Florestais. Lisboa. 23 pp.

Tomé, M., Ribeiro, F. 2000, Soares, P. O modelo GLOBULUS 2.1. Relatórios técnico-científicos do GIMREF nº 1/2001, Centro de Estudos Florestais, Instituto Superior de Agronomia, Lisboa, Portugal.

Tomé, M., Ribeiro, F. 2000. GLOBULUS 2.0, um modelo de aplicação nacional para a simulação da produção e crescimento do eucalipto em Portugal. Relatórios técnico-científicos do GIMREF nº 1/2000, Centro de Estudos Florestais, Instituto Superior de Agronomia, Lisboa, Portugal.

Tomé, M., Tomé, J., Ribeiro, F., Faias, S. 2007. Equações de volume total, volume percentual e de perfil do tronco para *Eucalyptus globulus* Labill. em Portugal. Silva Lusitana 15(1): 25-40.

Valentine, H.T., Gregoire, T.G., Burkhart, H, Hollinger, D.Y. 1997. A stand-level model of carbon allocation and growth, calibrated for loblolly pine. Canadian Journal of forest Research, 27: 817-830.

Valentine, H.T., Herman, D.A., Gove, J.H., Hollinger, D.Y., Solomon, D.S. 2000. Initializing a model stand for process-based projection. Tree Physiology. 20: 393–398.

Van Oijen, M, Rougier, J.C., Smith, R. 2005. Bayesian calibration of process-based forest model: bridging the gap between models and data. Tree Physiology, 25: 915-927.

Vanclay, J.K. 1994. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests. CAB International, Wallingford, UK.

Vega-Nieva, D.J., Tomé, M., Tomé, J., Fontes, L., Soares, P., Ortiz, L., Basurco, F., Rodrígez-Soalleiro, R. 2013. Developing a general method for the estimation of the fertility rating parameter of the 3-PG model: application in *Eucalyptus globulus* plantations in northwestern Spain. Canadian Journal of Forest Research, 43: 627-636.

Waring R.H, 2000. A process model analysis of environmental limitations on growth of Sitka spruce plantations in Great Britain. Forestry, 73: 65–79

Waring, R.H., Landsberg, J.J. Williams, M. 1998. Net primary production of forests: a constant fraction of gross primary production?, Tree Physiol. 18:. 129-134.

Waring, R.H., McDowell, N., 2002. Use of a physiological process model with forestry yield tables to set limits on annual carbon balances. Tree Physiology. 22:, 179–188.

Whitehead, D., Hall, G.M.J., Walcroft, A.S., Brown, K.J., Landsberg, J.J., Tissue, D.T., Turnbull, M.H., Griffin, K.L., Schuster, W.S.F., Carswell, F.E., Trotter, C.M., James, I.L., Norton, D.A. 2002 Analysis of the growth of rimu (Dacrydium cupressinum) in south westland, New Zealand, using process-based simulation models. International Journal of Biometeorology. 46: 66–75.

Xenakis, G., Ray, D., Mencuccini, M. 2008. Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. Ecological Modelling 219(1-2): 1-16

Chapter 1 - A system of compatible models to predict and project tree crown ratio and crown length for eucalypt plantations

Executive summary

A system of compatible equations to predict and project tree crown ratio and crown length in high-forest and coppice eucalyptus stands in Portugal was fitted to a large data set covering a wide range of age classes, locations and stand characteristics.

Abstract

- Context: Tree growth is strongly correlated with the crown size which is a good indicator of the tree´s vigour. Therefore, crown dimensions are important components of forest growth and yield models, namely individual tree models. Among the crown variables, crown length and crown ratio are often used as predictors.

- Aims: The objective of the present study is the development of a system of equations for the compatible prediction and projection of crown length and crown ratio in *Eucalyptus globulus* trees.

- Methods: The data used came from a large database that covers a wide range of situations and that includes series of measurements along time. After selecting the Richards function among other 0-1 bounded functions, seemingly unrelated regression was used to fit a system of three equations: crown ratio and crown length prediction and crown length projection.

- Results: The Richards function with asymptote equal to 1 was selected among a set of alternative functions. Functions to predict crown length and to project both variables were deducted from the crown ratio prediction model and the resulting system of compatible equations was fitted using non-linear seemingly unrelated regression.

- Conclusion: Parameters of the equations are a function of diameter at breast height, dominant height, stand density and age.

Keywords: crown ratio, crown length, compatible equations, ITSUR, *Eucalyptus globulus*

Introduction

Tree growth is strongly correlated with the crown size, which is a good indicator of the tree´s vigour. Therefore, crown dimensions can be important components of forest growth and yield models, namely of individual tree models. Crown ratio is often used as regressor in diameter at breast height (dbh) growth models (e.g. Soares and Tomé, 2003) and is an important variable for crown biomass estimation. The existing model to estimate *Eucalyptus globulus* aboveground biomass and biomass components (António *et al*. 2007) uses crown length as predictor for leaf and branches biomass.

Crown length (cl), defined as the difference between total tree height (h) and the height to the crown base (hc) is, in practice, difficult to measure, as the definition of the base of the live crown is subjective and imprecise, especially on asymmetric crowns. Measuring crown variables is not only time consuming, but also difficult to obtain in very dense stands and in very large trees, which can explain the relatively little research done on modelling crown parameters.

Several models for predicting crown ratio (cr) have been developed, using functions that vary between 0 and 1 such as the logistic (Ritchie and Hann 1987; Wykoff *et al.* 1981; Hasenauer and Monserud 1996; Temesgen *et al.* 2005) or the exponential (Belcher *et al.* 1982; Holdaway 1986, Dyer and Burkhart 1987, Hynynen 1995) functions. The analysis of these works shows that the variables related to crown length (expressed by cr, cl or hc) have been predicted as a function of tree and stand variables that reflect: 1) tree size, expressed by diameter at breast height (d), tree basal area (g) or tree slenderness (ts); 2) stand level competition through variables such as stand basal area (G), number of trees per ha (N) or the mean distance between trees (mdist); 3) hierarchical position of the tree in the stand or inter-tree competition evaluated by distance independent competition measures such as diameter relative to quadratic mean diameter or basal area of larger trees; 4) site quality (S).

By definition, crown ratio is bounded so that the values range between 0 (no crown) and 1 (full crown). This is the reason why the logistic function has been widely used on crown ratio predictions (e.g. Sprinz and Burkhart 1987; Soares and Tomé 2001). The logistic function was first introduced by Verhulst in 1945 for forest growth and used many times to model growth on biology. When the asymptote is set to 1 it is appropriate to model crown ratio. However, any function that is bounded between 0 and 1 can be used for this purpose, such as, for instance, probability density functions or growth functions with asymptote equal to one. Previous work from Soares and Tomé (2001), who developed a crown ratio equation for high-forest *Eucalyptus* plantations in Portugal, with a much smaller data set than the one used in the

present research, found that the Richards function, a generalization of the logistic growth function, shows a better performance than the logistic growth curve to model crown ratio.

The objective of the present study was the development of a system of equations for the compatible prediction of crown ratio and crown length in *Eucalyptus globulus* trees, as well as an equation to project crown length values over time with a self-referencing function.

Several models for predicting either crown ratio, height to the crown base or bole ratio have been developed, but only Dyer and Burkhart (1987) developed a system of compatible crown ratio and crown length equations. Valentine *et al.* (1994) derived a crown-rise model that directly estimates the average height to the base of the live crown, but no crown length or crown length projection model was ever developed.

The proposed system of equations was compared with the existing cr model (Soares and Tomé 2001) to asses if the new methodology represents a real improvement.

Material and Methods

II.1. Data

Eucalyptus globulus Labill. was introduced in Portugal 150 years ago for ornamental purposes but its fast growing ability was soon recognized by the pulp industry, making it the raw material for the Portuguese pulp and paper industry. The trees are planted at final density, as thinning and pruning practices are not usual during the first cutting cycle. The stands are intensively managed as a short rotation coppice system in which the first cycle of planted seedlings (single stem) is followed by 2 or 3 coppiced stands, with an average cutting cycle of 10–12 years (Soares and Tomé 2001) .

Data for this study came from the large database available in Portugal for eucalypt plantations. It includes data from permanent plots, continuous forest inventory, spacing trials, and data from an experiment to study the species potential productivity that includes fertilization and irrigation as treatments. Data from the irrigated stands were not used in the present study. From all the available data, only those where the height to the live crown base had been measured were selected. The base of the live crown was defined as the point of insertion of branches in at least three of the four horizontal quadrants defined around the stem of the tree. This measurement was taken mainly in live standing trees with a hypsometer, although some were made in felled trees. The available data set, after editing for deletion of abnormal points, included a total of 30 679 tree measurements.

First, as one of the objectives was the development of a model to project crown length, all the plots with only one measurement in time were set aside. Second, all the data pairs with

consecutive measurements apart from each other more than one year time plus or minus 1.5 months were also discarded. This way all the data in the fitting sub-set had at least two consecutive measurements with one year time interval (± 1,5 months) between them. After this procedure, the fitting sub-set reduced to 27 516 measurements covering a large range of site index, stand age and stand density (Table 1).

Table 1 – Characterization of the data from the fitting sub-set (n=27 516)

	minimum	mean	maximum
age (years)	0.5	5.3	12.5
d (cm)	0.2	9.0	32.9
h (m)	1.4	12	32.5
hbc (m)	0	6.5	26.0
cl (m)	0.01	5.5	15.3
cr	0.01	0.5	1.0
stand density (ha $^{-1}$)	586	2245	8563
S (m)	12.6	22.4	29.8

II.2. Model formulation

On the first stage of the model formulation, the 4 functions used by Soares and Tomé (2001) shown in Table 2 (all of them bounded between 0 and 1 if the restrictions of the parameters listed in Table 1 are fulfilled) were tested in order to confirm the superiority of the Richards function selected on that work. At this stage, the function f(d) selected by Soares and Tomé (2001), with k0 expressed as a linear function of the inverse of age (1/t), stand density (N) and dominant height (hdom) was used. The estimation of the parameters was made using the PROC NLIN procedure of SAS (SAS Institute Inc. 2009b) and the evaluation of the functions was made considering the SSE (Sum of Squared Errors), MSE (Mean Squared Error) and adjusted-R^2.

After selecting the function with the best results, several other tree and stand variables were tested to express the f(d) parameters, trying to improve the prediction ability of the model: tree age (t), dominant height (hdom), dominant diameter (ddom), diameter at breast height (d), site index (S), stand density (the scaling factor N/1000 was used instead of N in order to keep the coefficient estimates on the same scale (Schabenberger and Pierce 2002)), stand basal area

(G), the inverse of all these variables, tree slenderness (h/d), taper (d/h) and some competition indices, such as the ratio d/quadratic mean diameter. The possibility of f(d) = kd, with k as a linear function of the other tree and stand variables, was also tested, alongside varying values of the m and w parameters of the Weibull and Richards functions. To avoid a complex model, no more than 5 variables were tested at the same time. Once again, the PROC NLIN procedure of SAS (SAS Institute Inc. 2009b) and the non-linear least squares method were used and the final equation was selected considering the biological consistency of the parameters estimates, the proportion of variance explained by the model, SSE and MSE.

Table 2 – Functions tested for the crown ratio model

	Classic equation	Restrictions	Function used
exponential	$Y = A\,[1 - c\,e^{-kX}]$	A, c =1; kX>1	$cr = 1 - e^{-f(d)}$
logistic	$Y = \frac{A}{1 + c\,e^{-kX}}$	A, c =1; kX>1	$cr = \frac{1}{1 + e^{-f(d)}}$
Richards	$Y = \frac{A}{[1 + c\,e^{-k\,X}]^{1/m}}$	A, c =1; kX>1	$cr = \frac{1}{[1 + e^{-f(d)}]^{1/m}}$
Weibull	$Y = A\,[\,1 - c\,e^{-k\,X^w}]$	A, c =1; kX>1;w>0	$cr = 1 - e^{-f(d)^w}$

Note: f(d) - linear function of tree diameter k0+k1d with k0 and k1 expressed as linear functions of tree and stand variables; A - asymptote; c, k, w and m – function parameters

The crown length model was developed by multiplying the crown ratio model by total tree height, a rearrangement of the cr model to ensure numerically consistent estimates when using both models. After fitting the crown ratio equation, the crown length model was fitted to analyse the function behaviour using the form:

[1] $cl = h\,cr$

The availability of repeated measurements on the same tree made it also possible to fit a projection equation. The equation was developed by writing the expressions of crown length at times t-1 and t, calculating the ratio of the two expressions and manipulating the resulting expression in order to express crown length at time t as a function of its value at t-1 and tree and stand variables at the two points in time:

[2] $cl_t = f\,(cl_{t-1}, t-1, tree\ and\ stand\ variables)$

Age has been selected as an important predictor in the system of equations for crown ratio and crown length prediction. However, age is a variable that is not obtained as part of current

national inventory due to the non-existence of visible annual growth rings in eucalypt. Therefore a set of models without this variable, more easy to use in the cases where age is not available, was also developed.

II.3. Model fitting – OLS versus mixed-modelling approach

The data used in this study, repeatedly measured data from trees within plots, presents a multilevel hierarchical structure where various within- and between-plot correlations and heterogeneous variances occur. In addition, the correlations are both temporal and spatial. Violation of the independent and identically distributed (iid) error assumption could have important statistical consequences for analyses based on the least squares principle, invalidating hypothesis testing and interval estimation. The problem may be solved using mixed-modelling techniques (Davidian and Giltinan 1995).

While a number of studies have assumed that the fixed-effects component of a mixed-model represents the mean response or mean pattern of variation of the variable at population level, others have shown through model validation that this assumption may lead to poorer predictions than those obtained using a corresponding model fitted by least squares techniques (Meng and Huang 2010, de-Miguel *et al.* 2013). The advantage brought by a mixed model also comes at the price of added complexity, both statistical and computational. The selection between using OLS and mixed-models depends therefore on the objective of the model fitting. When the main objective is prediction with the mean response, without the possibility of calibration, OLS is preferable. In the present work, and since the objective is only to develop a system of equations to be used in a growth and yield model, the decision was not to use the mixed effects approach.

II.4. Testing and overcoming violations of the regression assumptions

The heteroscedasticity of the errors for each one of the three equations was analysed observing the plots of the studentized residuals versus the predicted values. In case of non-homogeneous variance of the errors, weighted regression was used. The normality of the residuals was analysed with the help of the normal QQ plots and corrected, when necessary, with robust regression using the Huber function (Myers 1986). The multicollinearity was evaluated by the condition number of the correlation matrix of the partial derivatives of the model in order to each parameter, computed at the solution. When this value is higher than 30, the collinearity may affect the variance of the estimates and for values of 100 the collinearity is considered very high (Belsley *et al.* 1980). According to Myers (1986), one must expect problems with multicollinearity when the condition number of the correlation matrix is higher

than the squared root of 1000, leading to instability of the regression coefficients estimates for higher values.

The use of repeated measurements from the same tree suggested the presence of auto-correlation among the model errors. Using the PROC MODEL (SAS Institute Inc. 2009a) procedure and the ODS (SAS Institute Inc. 2009c) tool of SAS, the plots of the autocorrelation function (acf) and partial autocorrelation function (pacf) were analysed in order to decide which structure best fitted the data: AR, MA or ARMA and the respective order.

II.5. Simultaneous fitting of the final system of equations

Our objective was the development of a system of compatible equations to predict crown ration and crown length and to project crown length from time t1 to time t2. Therefore, after the individual adjustments of each component of the system of equations, a simultaneous fitting was made to ensure the model compatibility. The functions and restrictions used ensure that no abnormalities arise and that the predicted values are logical (cr predictions are constrained between 0 and 1 and cl is less than total tree height). To have the best predictive ability from all the equations besides the numerically consistent estimates, simultaneous fitting is needed. Since the set of equations has contemporaneous cross-equation error correlation (also known as nonlinear seemingly unrelated regression system), the fitting was made using PROC MODEL (SAS Institute Inc. 2009a) and the ITSUR option, which provides an iterative estimation of the cross-equation covariance matrix. The initial values for the parameters and weight values used in the simultaneous fitting were the ones obtained in the individual fitting of each of the three equations of the system, cr, cl and cl projection. A requirement for model selection was the convergence of the simultaneous parameter estimation and the good performance of the model after the simultaneous fitting.

II.6. Model evaluation

The models developed in this study where evaluated considering:

- logic of model structure and biological aspects,
- visual or graphical inspection of observed data and residuals over model predictions and of QQ-plots
- cross validation procedure with PRESS residuals (residuals computed by a jack-knife procedure (Myers 1986)).

The following statistics based on the PRESS residuals were used:

- model fitting capacity (SSE, MSE and adjusted coefficient of determination);

- model bias (mean value of the PRESS residuals, MPRESS);
- model precision (mean absolute value of the PRESS residuals, MAPRESS).

Results

All models tested for the crown ratio prediction performed well, showing small differences between them, but the Weibull and the Richards functions originated smaller values for the SSE and MSE (Table 3). Looking at the statistics computed with the PRESS residuals, the Richards function is the one that is less biased.

Table 3 – Evaluation of the functions tested to model crown ratio

		Weibull		Richards
	exponential	w=10	logistic	m=6
$R^2_{adjusted}$	0.7473	0.7733	0.7676	0.7705
SSE	318.7	286.0	293.2	289.5
MSE	0.0116	0.0104	0.0107	0.0105
MPRESS	0.0028066	0.000971533	0.0020195	0.000078584
MAPRESS	0.0841997	0.0815008	0.0823681	0.0821165

Note: SSE: Sum of squared errors; MSE: Mean squared error; MPRESS: mean value of PRESS residuals; MAPRESS: mean absolute value of PRESS residuals

When comparing the results for crown ratio with the work of Soares and Tomé (2010), it can be seen that the values obtained for these statistics are similar. In that work, the logistic equation was the one with the higher value of adjusted-R^2 (0.76), followed by the Richards and Weibull functions (adj-R^2=0.77) and the exponential (adj-R^2=0.74).

Since the Weibull and the Richards functions performed both well, it was decided to make further tests using both functions. A first analysis of the Weibull function showed that not only the best predictor variables to be used in the function f(d) were t, N, hdom and d, the ones from Soares and Tomé (2001), but also that the best value for the w parameter was 10, again the same used in that work. The same analysis made on the Richards function showed the same predictor variables, but the best fit was obtained with m=20 instead of the value 6 obtained by Soares and Tomé (2001). In both cases, using f(d)=k0+k1d and making k0 a

function of all the parameters gave better results than using f(d)=kd and k as a linear function of the parameters.

An evaluation of the final models selected for each function is shown on Table 4 and the QQ-plots of the residuals and the plots of actual versus predicted values of cr on Figure 1.

Table 4 – Nonlinear ITSUR summary of residual errors for the crown ratio model

	SSE	MSE	R^2_{adj}
Weibull with w=10	286.0	0.0104	0.7733
Richards with m=20	288.1	0.0105	0.7716

Note: SSE: Sum of squared errors; MSE: Mean squared error

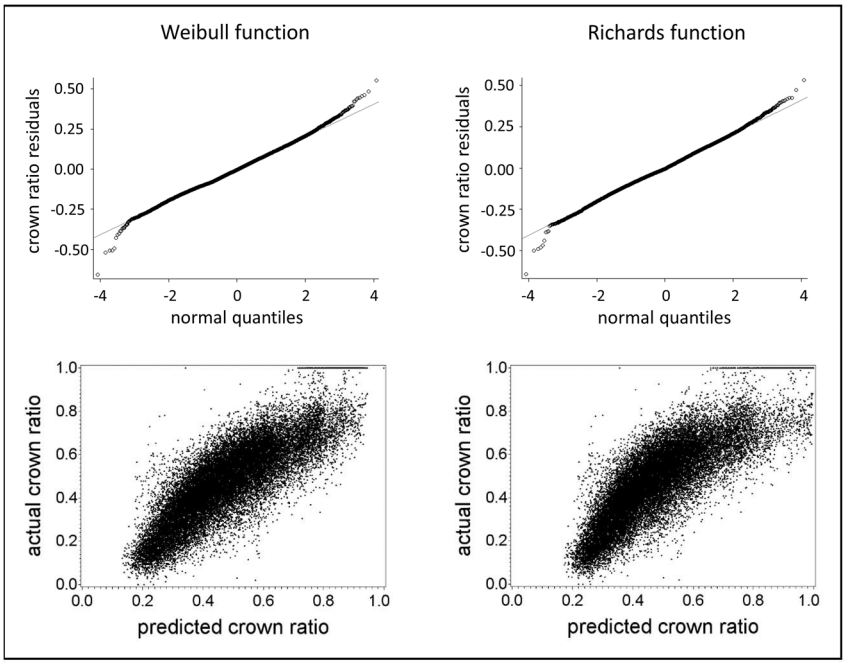


Figure 1. QQ plot and graphics of the actual over predicted values of cr for the Weibull and Richards function, respectively on the left and on the right.

The Weibull function presents slightly better values for the evaluation statistics than the Richards function and has also the ability to better predict small crown ratio values available in the database, which is an important characteristic for the older trees with smaller crown dimensions. The goodness-of-fit statistics for the Richards function are very similar to the Weibull ones and this last function is not often used in crown ratio modelling, so both functions

were tested for the fitting of the final system of equations. In this stage the system of equations based on the Weibull function had convergence problems that could not be overcome, so the system of equations finally selected to predict crown ratio and crown length was the one based on the Richards function, the same function selected by Soares and Tomé (2001).

The selected crown ratio equation was:

[3] $$cr = \frac{1}{\left[1+e^{-f(d)}\right]^{1/m}}$$

with $f(d) = k0 + k1\,d$, $k0 = a1 + a2\left(\frac{1}{t}\right) + a3\left(\frac{N}{1000}\right) + a4\,hdom$ and m=20

The analysis of the autocorrelation plots showed that the autocorrelation function (acf) rapidly decreases to zero and that the partial autocorrelation function (pacf) abruptly cuts after lag 1, the typical patterns for a first-order autoregressive model - AR (1) - for the autocorrelation structure of the errors. The value close to zero of the plotted pacf at lag 1 means that the AR (1) model removes all the autocorrelation. The AR process expresses the series in terms of past observations and the current disturbances (random error), so this error structure makes sense.

The QQ plots showed evidence of non-normal distribution of the model errors, so a correction was made using robust estimation with Huber´s function (Myers 1986).

The final QQ plot and plot of the studentized residuals over predicted values for cl are shown, as an example, on figure 2. The analysis of the condition number of the correlation matrix showed no problems of multicollinearity, since all the values were less than 30.

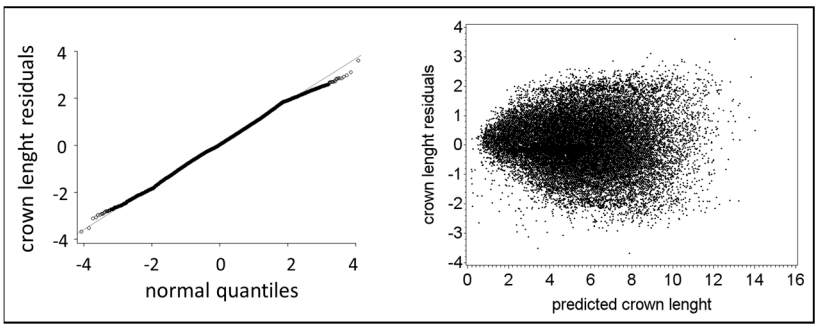


Figure 2. QQ plot and graphic of the studentized residuals over predicted values of crown length (m) for the Richards function.

The crown length equation was adjusted and evaluated using the form:

[4] $$cl = h\ \frac{1}{\left[1+e^{-\left[a1+a2\left(\frac{1}{t}\right)+a3\left(\frac{N}{1000}\right)+a4\,hdom+a5\,d\right]}\right]^{1/20}}$$

For the crown length projection, the equation used was

$$[5] \qquad cl = cl_{t-1}\left(\frac{h}{h_{t-1}}\right)\left(\frac{\left[1+e^{-[a1+a2\left(\frac{1}{t-1}\right)+a3\left(\frac{N_{t-1}}{1000}\right)+a4\,hdom_{t-1}+a5\,d_{t-1}]}\right]}{\left[1+e^{-[a1+a2\left(\frac{1}{t}\right)+a3\left(\frac{N}{1000}\right)+a4\,hdom+a5\,d]}\right]}\right)^{1/20}$$

After individually fitting all the equations (crown ratio, crown length and crown length projection), a similar set of models, but without the age variable was fitted. The tests showed that including this variable, the model performed slightly better, with better values of the evaluation statistics (Table 5), but the differences are small, which means that even if age is not available, a good prediction of the crown ratio and crown length is still possible.

Table 5 – Nonlinear OLS summary of residual errors for the 3 models (cr, cl projection, cl), individually fitted, with and without age

		Degrees of freedom					
		Model	Error	SSE	MSE	Root MSE	R^2_{adj}
with age	cr	6	22154	184.3	0.00832	0.0912	0.8180
	cl projection	5	17551	19234.9	1. 0959	1.0469	0.8217
	cl	6	22154	20580.7	0.9290	0.9638	0.8428
without age	cr	5	22155	242.1	0.0109	0.1045	0.7609
	cl projection	4	17552	20222.5	1.1521	1.0734	0.8126
	cl	5	22155	22260.8	1.0048	1.0024	0.8300

Note: SSE: Sum of squared errors; MSE: Mean squared error

The analysis of the fitting statistics (Tables 6 and 7) showed that, after the simultaneous fitting, the equations present good prediction ability, with adjusted R-squared values around 80 percent. Relatively to the individual fitting (Table 5), all the models presented a slight decrease in the model fit, but nothing too relevant. The simultaneous fitting with the ITSUR imposes restrictions on the parameters estimates, so this was already expected.

Table 6 – Nonlinear ITSUR summary of residual errors for the 3 models (cr, cl projection, cl) with age

	Degrees of freedom					
	Model	Error	SSE	MSE	Root MSE	R^2_{adj}
cr	2.667	17553	152.1	0.00866	0.0931	0.8148
cl projection	2. 667	17553	20580.2	1.1724	1.0828	0.8093
cl	2. 667	17553	17838.6	1.0163	1.0081	0.8112

Note: SSE: Sum of squared errors; MSE: Mean squared error

Table 7 – Nonlinear ITSUR Summary of residual errors for the 3 models (cr, cl projection, cl) without age

	Degrees of freedom					
	Model	Error	SSE	MSE	Root MSE	R^2_{adj}
cr	2.333	17554	193.1	0.0110	0.1049	0.7649
cl projection	2.333	17554	22565.8	1.2855	1.1338	0.7909
cl	2.333	17554	18511.3	1.0546	1.0269	0.8041

Note: SSE: Sum of squared errors; MSE: Mean squared error

The plots of the observed versus predicted values show a close to linear relationship for both systems of equations, exception noted for the crown ratio, for which the models could not predict such small and high values as some of the observed. The plot closely follows the 1:1 line, with some dispersion, more evident in the case of the crown ratio functions (Figure 3).

The validation shows that both crown ratio (Figure 4) and crown length (Figure 5) models present a small bias, more evident in the plots against age classes and for the crown ratio equations, showing positive bias for young stands (age under 4 years), no bias for stands in the age classes 4 to 8 years and 8 to 12 years and again a positive bias for the stands older than 12 years, but nothing that indicates serious bias in the models. From figure 3 is was already visible that the equations for crown ratio where slightly underestimating both the smaller and the bigger values present in the data set, that correspond to young and the old trees.

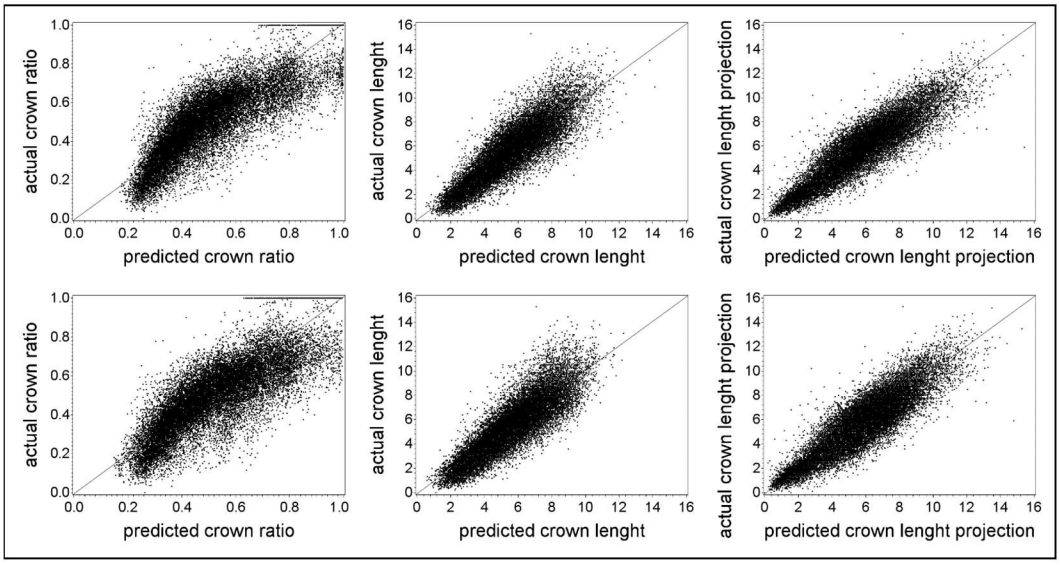


Figure 3. Plots of observed versus predicted values for crown ratio, crown length (m) and crown length projection (m) for the models with age (above) and without this variable (below).

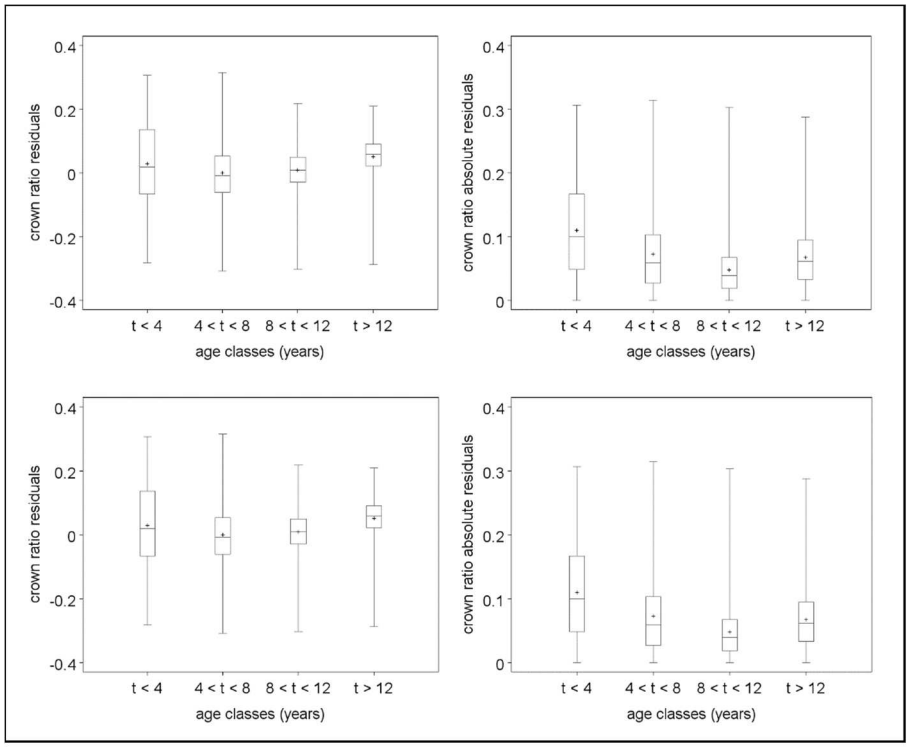


Figure 4. Box-plot graphics of prediction residuals against age classes for crown ratio in the model with age and without age, respectively above and below. The bottom and top edges are defined by Q1 and Q3. The marker inside the box indicates the mean value. The line inside the box represents the median value. The whiskers extend to the maximum and minimum values, excluding outliers.

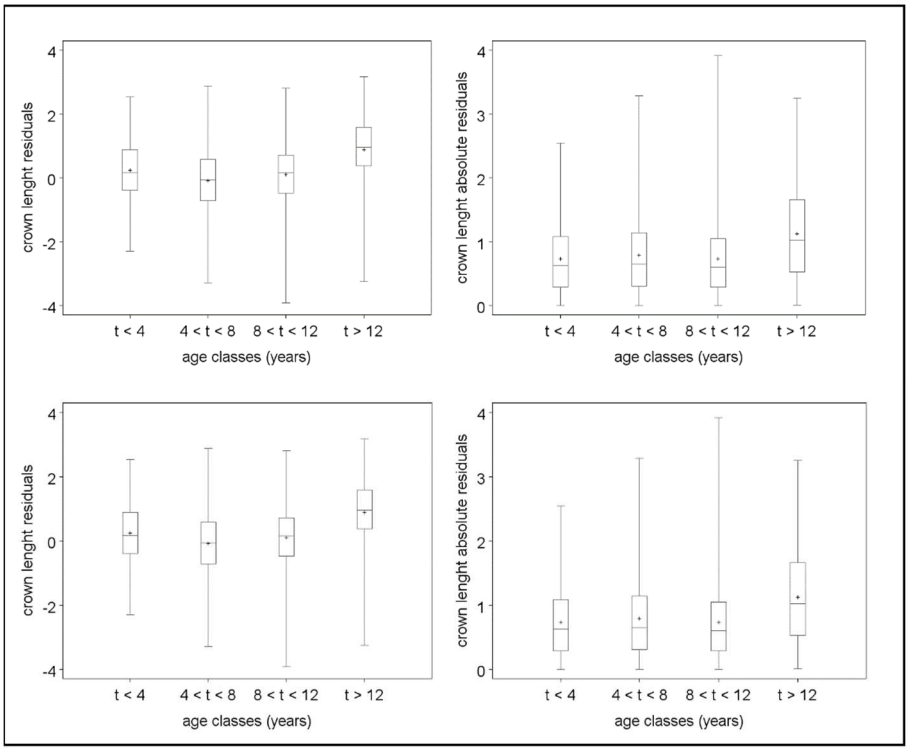


Figure 5. Box-plot graphics of prediction residuals against age classes for crown length (m) in the model with age and without age, respectively above and below. The bottom and top edges are defined by Q1 and Q3. The marker inside the box indicates the mean value. The line inside the box represents the median value. The whiskers extend to the maximum and minimum values, excluding outliers.

The crown length projection model presents no bias or precision problems, as it can be seen in Figure 6.

After the simultaneous fitting, the final models are:

[6] $cr = \frac{1}{\left[1+e^{-f(d)}\right]^{1/20}}$,

[7] $cl = h\,\frac{1}{\left[1+e^{-f(d)}\right]^{1/20}}$ and

[8] $cl = cl_{t-1}\left(\frac{h}{h_{t-1}}\right)\left(\frac{\left[1+e^{-f(d\,t-1)}\right]}{\left[1+e^{-f(d\,t)}\right]}\right)^{1/20}$,

with:

1. model with the age variable

$$f(d_i) = -8.93108 + 25.52811\left(\frac{1}{t_i}\right) - 1.35096\left(\frac{N_i}{1000}\right) - 1.00667\,hdom_i + 0.627747\,d_i$$

2. model without the age variable

$$f(d_i) = 7.652947 - 1.09608\left(\frac{N_i}{1000}\right) - 1.63921\, hdom_i + 0.571169\, d_i$$

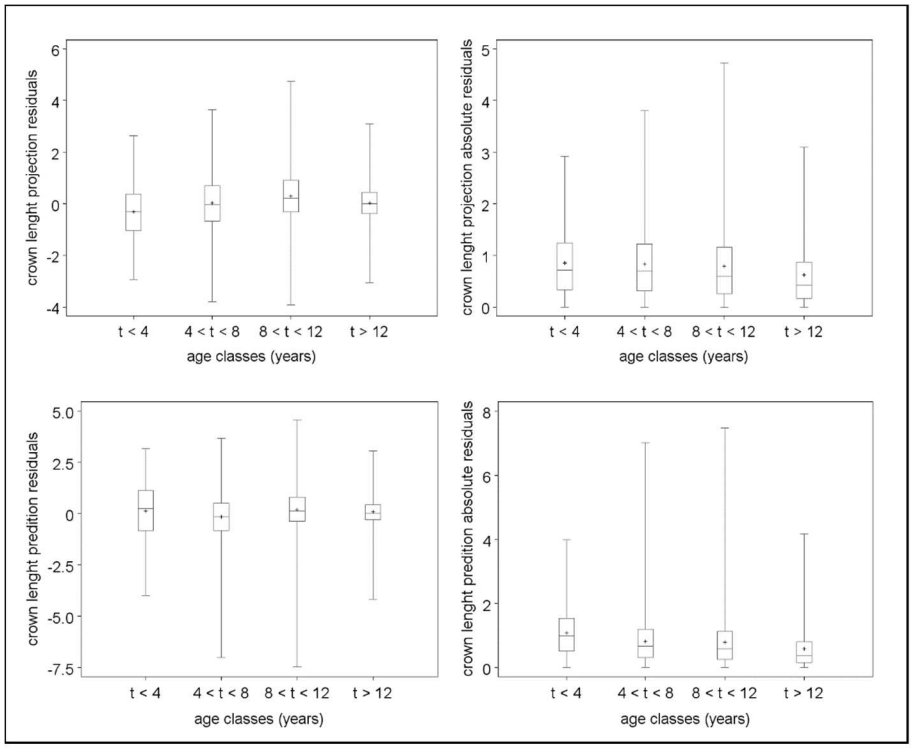


Figure 6. Box-plot graphics of prediction residuals against age classes for crown length projection (m) in the model with age and without age, respectively above and below. The bottom and top edges are defined by Q1 and Q3. The marker inside the box indicates the mean value. The line inside the box represents the median value. The whiskers extend to the maximum and minimum values, excluding outliers.

Discussion

Crown ratio is a very effective predictor variable in many growth and yield relationships, so there has been some effort in its modelling. There are numerous models for predicting crown dimensions, from simple relationships between the average height to the base of the live crown and the average tree height to more complex models including the number of trees per unit area (N) or the average inter-tree distance (mdist) (Valentine *et al.* 1994). Several tree and stand variables have been related to height to the crown base, crown ratio (Wykoff *et al.* 1981; Belcher *et al.* 1982; Hasenauer and Monserud 1996; Soares and Tomé 2001; Temesgen *et al.* 2005; Leites *et al.* 2009) and crown height (Dyer and Burkhart 1987). The most usual tree and stand variables used as regressors are total tree height, diameter at breast height, tree

slenderness (ts, given by the ratio total tree height/diameter at breast height), age, stand basal area, site index, crown competition factor, tree competition measures and site factors.

Logistic models are used more often in crown dimensions modelling than exponential ones, because although exponential ones are simple and may present meaningful parameters, they can also predict cr values greater than 100% in the extreme of the data. The use of logistic equations or other growth functions is advantageous because predictions can be easily constrained between 0 and 1. The Weibull function was only found referenced once in crown ratio modelling (Dixon 1985), so it was expected that the logistic or the Richards function where the ones with better performance from the four tested.

Soares and Tomé (2001) developed a model to predict crown ratio for high-forest eucalyptus plantations using data from the north and central coastal regions of Portugal. In this work data from other regions and also coppice stands were used in order to obtain equations applicable to the whole area of distribution of the *Eucalyptus globulus* in Portugal. The other limitation of the previous equation was the range of ages present in the data available at the time the equation was developed. There was a lack of data from young stands and from stands older than 8 years, which limited the model application. Additionally, the system of equations presented in the present work includes not only a crown ratio model (as was the case in Soares and Tomé 2001), but also equations to predict and project crown length that are compatible with the crown ratio prediction. The model selected for crown ratio in both Soares and Tomé (2001) and the present work are the same, but the parameter values are different, the present model leading to more accurate predictions. The present work also developed a second system of equations that does not need stand age, a stand level variable that is not always accessible and that can constitute an additional problem in uneven-aged stands.

Figure 7 presents the evaluation of the mean residuals, per age classes, of the cr equation developed versus the cr equation from Soares and Tomé (2001). In the first age class, t under 4 years, the proposed equation performs better than the existing one, both in bias and precision. The advantage of the new model is even clearer in all age classes over 6 years. In the age class of 4 to 6 years the cr equation from Soares and Tomé (2001) presents slightly better values for bias and precision as most the data used in that work was between 4 and 8 years old, so it was expected that the equation had a good result for this age interval. The present work aimed at improving the equation in the younger age classes where the differences between predicted and measured values was very big and on the older age classes for which there were no data available at the time Soares and Tomé (2001) developed their equation.

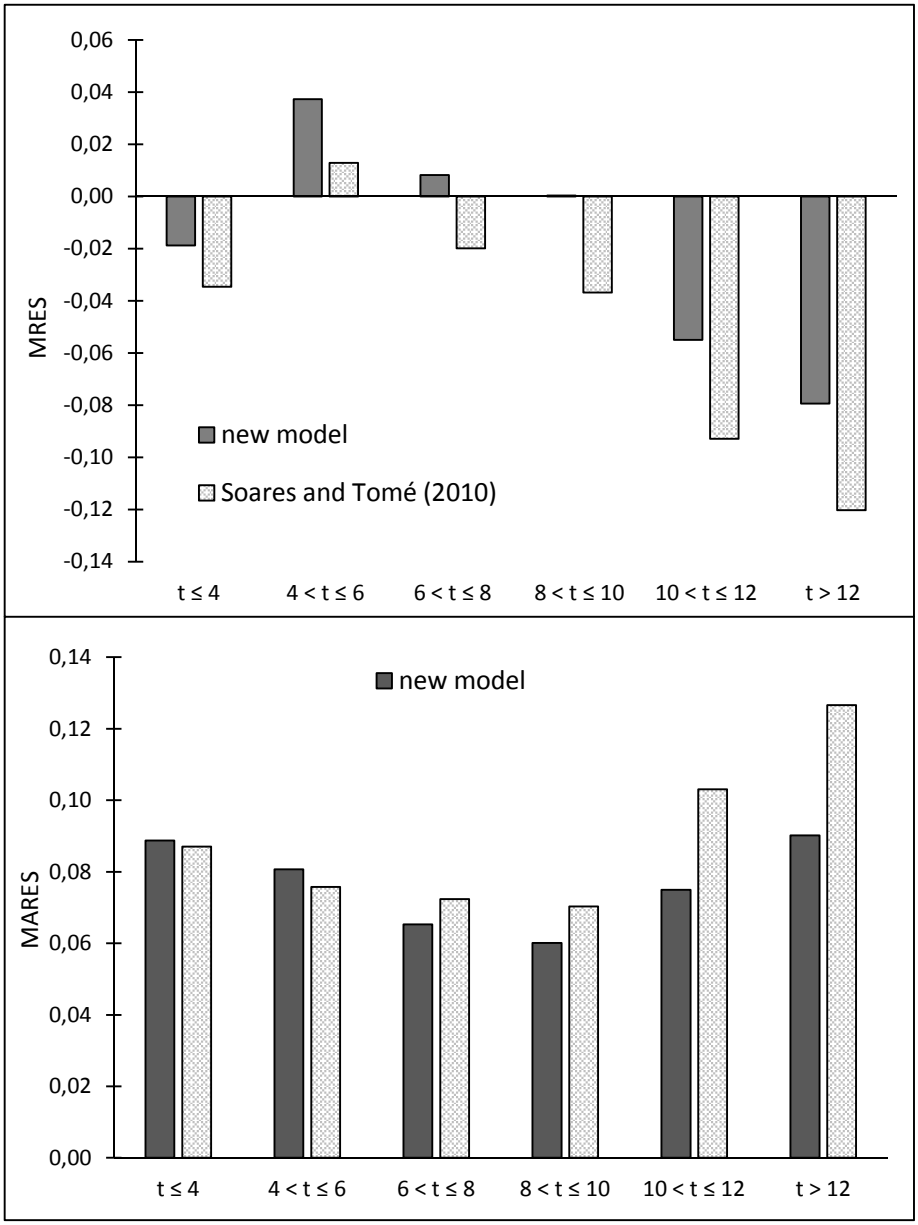


Figure 7. Column charts of the mean prediction residuals (MRES) and absolute prediction residuals (MARES) against age classes for crown ratio and the crown ratio equation from Soares and Tomé (2010), respectively above and below.

The parameters estimates present logic signals and the models are robust and with good prediction ability. Age and dominant height are all inversely related to crown ratio, which makes sense from a biological point of view. Older and taller trees have smaller crown ratios. Density is inversely related to crown ratio while diameter at breast height has an associated positive parameter which indicates higher crown ratio values for higher trees. Within a stand, dominant trees show larger crowns. Individual tree diameter is not necessarily directly correlated with

crown ratio, but its influence on crown ratio depends on the characteristics of the stand in which the tree is located. Density alone is also not directly related to crown ratio. Spaced trees can have bigger or smaller crowns, depending on tree age, height and other characteristics. It is important for crown ratio prediction, but its influence is complex and cannot be easily translated into a parameter sign.

Conclusion

The objective of the present study was the development of a system of equations to predict, in a compatible way, crown ratio and crown length and to project crown length in *Eucalyptus globulus* plantations in Portugal.

As stand age was selected as one of the predictor variables and this variable is not always available from operational forest inventories, two system of equations, one including stand age as regressor and another without this variable were finally selected (equations [6] to [8]).

Other than the system of equations, some conclusions could also be drawn:

1. The non-symmetric functions, such as the Richards or the Weibul functions, perform better than the logistic function to model crown ratio and crown length.

2. The parameters from the crown ratio and crown length equations are different among stands expressing influence of different factors:

- The fact that crown ratio decreases as the stand develops is expressed by the negative value of the parameter associated with dominant height in both systems (with and without age). The system that includes age as regressor emphasizes this fact by adding a positive relationship with the inverse of age.

- The influence of stand density in decreasing crown ratio is expressed in both systems of equations by the negative sign associated with stand density.

- The differences in crown ratio within a stand, with larger crowns associated with larger trees, are expressed by the positive sign associated with tree diameter. The inclusion of a distance-independent competition index was not significant.

The two system of equations developed in the research presented here, based on a fitting data set covering a great range of situations, represents an important tool for generalized use to predict crown ratio and crown length for *Eucalyptus globulus* in Portugal and allows for more accurate predictions of aboveground biomass components using the existing model of equations from António *et al*. (2007).

The methodology presented can be applied to other species provided the data needed are available.

References

António N, Tomé M, Tomé J, Soares P, Fontes L (2007) Effects of tree, stand, and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research, 37, pp. 895–906

Belcher DW, Holdaway MR, Brand GJ (1982) A description of STEMS, the Stand and Tree Evaluation and Modelling System. General technical report NC 79. St. Paul, Minn., U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station. Available from http://nrs.fs.fed.us/pubs/143 [accessed 18 June 2009]

Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: Identifying Influential data and sources of collinearity, John Wiley & Sons. New York.

Cook RD (1994) An Introduction to Regression Graphics. John Wiley & Sons, New York, 253 pp.

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman & Hall, New York, 360 p.

de-Miguel S, Guzmán G, Pukkala T (2013) A comparison of fixed- and mixed-effects modelling in tree growth and yield prediction of an indigenous neotropical species (*Centrolobium tomentosum*) in a plantation system. Forest Ecology and Management, 291: 249–258.

Dixon GE (1985) Crown Ratio Modelling Using Stand Density Index and the Weibull Distribution. Internal Report. U.S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO, 13 pp.

Dyer ME, Burkhart HE (1987) Compatible crown ratio and crown height models. Canadian Journal of Forest Research. 17(6): 572-574.

Hasenauer H, Monserud R A (1996) A crown ratio model for Austrian forests. Forest Ecology and Management, 84(1-3): 49-60.

Holdaway M R (1986) Modelling tree crown ratio. The Forestry Chronicle, 10:451–455.

Hynynen J (1995) Predicting tree crown ratio for unthinned and thinned Scots pine stands. Canadian Journal of Forest Research 25(1):57-62.

Leites LP, Robinson AP, Crookston NL (2009) Accuracy and equivalence testing of crown ratio models and assessment of their impact on diameter growth and basal area increment predictions of two variants of the Forest Vegetation Simulator. Canadian Journal of Forest Research, 39: 655–665.

Meng SX, Huang S (2010) Incorporating correlated error structures into mixed forest growth models: prediction and inference implications. Canadian Journal of Forest Research, 40: 977-990.

Myers RH (1986) Classical and Modern Regression with Applications. Second edition. Duxbury Classic Series. 357 pp.

Ritchie MW, Hann DW (1987) Equations for predicting height to crown base for fourteen tree species in southwest Oregon. Research Paper 50, Forest Research Lab, Oregon State University, Corvallis, Oregon. Available from http://hdl.handle.net/1957/8246 [accessed 3 June 2009]

SAS Institute Inc. (2009a) SAS/ETS ® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 2876 p.

SAS Institute Inc. (2009b) SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc. 7886 p.

SAS Institute Inc. (2009c) SAS ® 9.2 Output Delivery System: User's Guide. Cary, NC: SAS Institute Inc. 984 p.

Schabenberger O, Pierce FJ (2002) Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press LLC 2000. Boca Raton, FL. pp. 211-213 and pp. 539-545. 738p.

Soares P, Tome M (2001) A tree crown ratio prediction equation for eucalypt plantations. Annals of Forest Science 58(2): 193-202.

Soares P, Tomé M (2003) GLOBTREE, an individual tree growth model for *Eucalyptus globulus* in Portugal. In: Amaro, A., Reed, D. E Soares, P. (Eds), Modelling forest systems. CAB International, pp.97-110.

Sprinz PT, Burkhart HE (1987) Relationships between tree crown, stem, and stand characteristics in unthinned loblolly-pine plantations. Canadian Journal of Forest Research, 17(6): 534-538.

Temesgen H, LeMay V, Mitchell SJ (2005) Tree crown ratio models for multi-species and multi-layered stands of south-eastern British Columbia. Forestry Chronicle 81(1): 133-141. Available from http://pubs.cif-ifc.org/doi/pdf/10.5558/tfc81133-1 [accessed 3 June 2009]

Valentine HT, Ludlow AR, Furnival GM (1994) Modelling crown rise in even-aged stands of sitka spruce or loblolly-pine. Forest Ecology and Management. 69(1-3): 189-197.

Wykoff W, Crookston NL, Stage AR (1981) User's guide to the stand prognosis model. General technical report INT 133. Ogden, Utah, U.S. Dept. of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. Available from http://www.fs.fed.us/rm/pubs_int/int_gtr133.pdf [accessed 18 June 2009].

Chapter 2 - Modelling the aboveground stand biomass of planted and coppiced *Eucalyptus globulus* in Portugal

Executive summary

Two systems of equations – one considering stand age and the other one without considering it - to predict stand level aboveground tree component biomass (stem, bark, branches and leaves) for high-forest and coppice eucalyptus stands in Portugal were fitted.

Abstract

- Context: The growing need for information on carbon stocks and carbon sequestration by forests has made forest biomass estimation a relevant issue, far more important in the case of fast growing and intensively managed species such as *Eucalyptus globulus* Labill. as they have a large potential to contribute for the national commitments in relation to carbon sequestration by forest resources.

- Aims: In this paper two systems of equations for predicting stand level total aboveground biomass and biomass per tree component (stem, bark, branches and leaves) taking into consideration the specific characteristics of the stand. The two systems differ by considering or not stand age as one of the regressors. Stand age leads to better stand level biomass estimations but it is not always available in forest inventory, therefore the need to develop a system without this variable. The systems are based on equations that use stand variables that usually come as result of forest inventories as regressors so that they can be easily used.

- Methods: The models where developed using the PROC MODEL procedure of SAS statistical software and, in order to account for the heteroscedasticity and non-normality of the model errors, weighted regression using the Huber function was used.

- Results: In a final stage, the chosen equations for both systems where simultaneously adjusted, using the non-linear seemingly unrelated estimation method applied iteratively, and considering that total aboveground estimates were equal to the sum of the biomass components estimates thus guarantying the property of additivity among the biomass components and the total aboveground biomass.

- Conclusion: The models have a good predictive capacity using only stand variables easily accessible in the forest inventories such as stand age, stand density, dominant height, basal area and cultural regime.

Key-words: *Eucalyptus globulus* Labill.; stand level biomass equations; allometric equations; forest inventory

Introduction

The growing need for information on carbon stocks and carbon sequestration by forests has made forest biomass estimation a relevant issue. Quantifying aboveground biomass is important not only to support management, but also as quantity assessment for the industry, carbon sink evaluation and analysis of nutrient cycles. This information becomes even more relevant in fast growing intensively managed, short rotation species such as *Eucalyptus globulus*.

Direct biomass measurements in trees is only viable at a small scale and even in this case it always implies the sampling of a small number of trees. Stand biomass is therefore usually estimated using field data or remote sensing techniques.

Remote sensing techniques have been drawing attention due to time and cost reduction when compared with specific data gathering in the field (Bettinger and Hayashi, 2006). These techniques have been having success, but as the forest structure gets more and more complex, the estimates start to lack reliability and more information at stand level is needed, not only to increase estimates reliability (Bettinger and Hayashi, 2006), but for calibration and validation (Schroeder *et al.*, 1997). Remote sensing biomass estimation is still a challenge, investigation is still needed to understand and identify the flaws in this process and reduce the estimates uncertainty (Lu, 2006). Field data is therefore still considered the most precise option for aboveground biomass estimations (Bettinger and Hayashi, 2006). The combination of forest inventory data with estimation models is the most used method to estimate stand biomass. The inventory data are collect at a large scale, usually include all the interest populations, the data collection method is made in a way that makes the data easily available and in quantity and therefore statistically valid (Schroeder *et al.*, 1997). Biomass estimation models may be developed at tree or stand level. The first are applied to each tree within each inventory plot and tree level biomass estimates are summed-up for the plot and expanded to ha while the second allow the direct estimation of stand biomass from other easily available stand variables such as stand basal area and dominant height. Tree level biomass models are much more common than stand level biomass estimations (e.g. Zianis *et al.*, 2005 compiled 607 equations for biomass estimation for tree species growing in Europe). However, stand level biomass models are an important component of whole stand growth and yield models and are needed to estimate stand biomass when pre-processed forest inventory data are available (for instance information from previous forest inventories).

Stand level biomass estimations may be obtained by biomass prediction models (e.g. Monserud *et al.*, 1996; Tomé *et al.*, 2001; Fang *et al.*, 2001; Tomé *et al.*, 2006; Gonzáles-García *et al.*, 2013) or through the so-called Biomass Expansion Factors (BEF´s) that directly

convert volume into biomass (e.g. Schroeder *et al.*, 1997, Brown *et al.*, 1999; Fang *et al.*, 2001). BEF's are defined as the reason between biomass and volume and estimate biomass by multiplying the BEF by stand volume. One problem with this methodology is the fact that they are based on a volume estimation which has also associated some error. BEF's vary between species, age, place, stand characteristics and also volume but it is very common to assume a species-specific value. Some authors (e.g. Soares *et al.*, 2012; Faias *et al.*, 2009; Teobaldelli *et al.*, 2009) developed equations to predict the BEF value taking into account other stand variables. Gonzáles-García *et al.* (2013) compared three approaches - biomass equations, BEF equations and constant BEF´s – to predict aboveground biomass at stand level in *Eucalyptus nitens* and concluded that assuming a constant BEF was the least accurate method for biomass estimation; BEF equations were slightly more accurate than biomass equations for the crown component estimation and biomass equations provided the most accurate predictions for the stem component and for total aboveground biomass. Considering that the use of BEF´s has the intermediate step of volume estimation with an associated error and that a BEF equation, if conveniently expressed, is just an alternative biomass equation that uses stand volume as one of the predictors, it may be concluded that biomass equations remain the best option to estimate stand level aboveground biomass components.

Monserud *et al.* (1996) made a literature review of stand level models for biomass components in Russia. Semechkina (1978) concluded that wood biomass can be predicted as a function of basal area; branch biomass by the stand medium diameter and needles biomass by a stand age function. Gorbatenko (1970) realized that basal area and stand volume where the best variables to predict wood, roots, needles and crown biomass. Kuzikov (1979) studied several linear and non-linear regional models to predict biomass as a function of stand age, height and basal area and Usol´tsev (1988) studied the biomass structure based on the Biomass/Volume ratio and found as predictors stand age, stand density and site index. Being stand age, site index and volume the most listed variables in the literature review, this where the ones used in Monserud´s biomass equations.

The use of allometric equations to estimate above-ground biomass in Australia (Snowdown *et al.*, 2000) showed evidence that application of 'stand-scale' equations is likely to be as effective as the use of more site- and species-specific equations applied to individual trees in a stand. For example, stand-based equations that enable biomass to be predicted from basal area, mean height, stocking, or combinations of these variables, appear to be reasonably robust (Snowdown *et al.*, 2000).

According to Clutter *et al.* (1983), explicit functions can be used to predict actual production in terms of biomass and volume using stand height, a stand density indicator, basal area, age and site index. Following this principle, Barrio-Anta *et al.* (2006) considered reasonable to

relate stand production with the product of biomass or volume of the representative tree – given by quadratic mean diameter and dominant height – and the number of trees per hectare and estimate stand biomass or volume in this way. This is very similar to the medium tree technique developed in the 60´s and 70´s by several investigators (Baskerville, 1965; Madgwick and Satoo, 1965; Attiwill and Ovington, 1968; Crown, 1971).

The growth and yield model Globulus (Tomé at al., 2001; Tomé *et al.*, 2006) includes a system of allometric equations for total biomass and biomass per tree components biomass estimation at stand level. The equations are site-specific by expressing the parameters as a function of dominant height, basal area, site index, stand density and age to localize the parameters.

The main objective of this study was to develop a system of compatible equations to estimate eucalyptus aboveground biomass, total and per tree component, at the stand level, expressed in terms of dry weight, from readily available stand and site variables. The system must provide a consistent basis for evaluating forest biomass across regional boundaries as well as to provide compatible predictions of stand level total aboveground biomass and of the largest aboveground biomass components of the tree: stem wood, stem bark, branches and leaves. The study must also contribute to give highlights to some unanswered questions that were also investigated in preparing this paper:

1. Is there a significant improvement in the predictive ability of the models when the allometric constants are expressed as a function of stand and site variables, such as age, dominant height, stand density, site index or climate variables?
2. Is there a need to use different parameter estimates for planted and coppice regenerated stands?

Material and methods

II.1. Data

The data used in this work came from the available database for eucalypt plantations in Portugal that includes data from permanent plots, continuous forest inventory, spacing and fertilization trials. The plots are located throughout the species area of distribution in Portugal and represent the existing range of ages, stand densities, sites and management options. The existing data from some permanent plots and trials also gives information from less usual management options, such as high densities and older trees, since the usual rotation cycle for eucalyptus doesn´t tend to go beyond 11/12 years. The plot size ranged from 100 to 2464 m^2, depending on the source of the data. Continuous forest inventory plot size is usually 400-500 m^2, but in permanent plots and trials, plot size varies. Diameter at breast height was measured

in every tree and total tree height in dominant, model trees or all the trees, depending on data provenience. Non-measured tree heights were estimated with the height-diameter curve from Tomé *et al.* (2007). Aboveground biomass and biomass per tree component (stem wood, stem bark, branches and leaves) were estimated using the system of biomass equations developed by António *et al.* (2007). Crown length, used as regressor in the leaf and branches biomass equations, was estimated with the crown ratio and crown length system of equations developed by Oliveira and Tomé (2015).

Stand variables for each plot included age (t, years), stand density (N, trees ha^{-1}), basal area (G, m^2 ha^{-1}), dominant height (hdom, m), site Index (S, m) and total aboveground biomass (Wa) and biomass per tree component (Wi, Mg ha^{-1}, i=w for stem wood, b for stem bark, br for branches and l for leaves).

The available data set, after editing for deletion of abnormal points, included a total of 22459 measurements (Table 1).

Table 1 – Characterization of the data used to fit the biomass equations (n=22459)

	minimum	mean	maximum
Age (years)	0.6	8.4	33.9
Stand density (ha^{-1})	67	1168	8563
Basal area (m^2 ha^{-1})	0.05	11.89	77.88
Dominant height (m)	4.68	16.81	36.8
Site index (m)	1.27	19.12	33.93
Stem wood biomass (Mg ha^{-1})	0.09	40.70	434.96
Stem bark biomass (Mg ha^{-1})	0.03	5.30	64.78
Leaves biomass (Mg ha^{-1})	0.11	4.88	20.11
Branches biomass (Mg ha^{-1})	0.07	4.76	24.41
Total aboveground biomass (Mg ha^{-1})	0.36	55.87	540.50

II.2. Methods

The objective of this work was to develop simple systems of equations to predict aboveground total biomass and biomass per tree component from easily obtained stand variables. It has

been shown (António *et al.*, 2007) that the allometric model provides accurate tree level biomass estimations in *E. globulus*. Tree dbh has been selected for biomass components, complemented by total tree height for stem wood and stem bark and by crown length for the crown biomass components. This gives an indication of the adequacy of the allometric model for stand level biomass prediction. The following formulation was used in this research:

$$W = a\, G^{b}\, hdom^{c}$$

where W – aboveground biomass or a biomass component; G – basal area; hdom – dominant height and a, b and c are parameters expressed as linear functions of stand level variables: age, site index and stand density. Since the data had information regarding stands originated from seedlings or by coppice, it was decided to test also a dummy variable for coppice with the value 0 in seedling stands and value 1 in coppice regenerated stands). Site index (S) is an indirect measure of the potential growth at a site that assumes that the dominant height growth of a stand is independent of changes in the environment and that it is not influenced by stand stocking. When growth conditions are stable, one can use this concept without problems, but when the conditions are variable, the growth response is highly non-linear and the relationships can be very difficult to predict. It is known, for example, that the relationship between tree height and diameter varies with altitude in such a way that trees that are grown in higher altitudes tend to be shorter and thicker (Turnbull *et al.*, 1993), resulting in different productivities for the same value of S. Considering this variability issues and also the fact that S is dependent on the equation used to estimate it, it was decided that the inclusion of S as a variable to localize the a, b and c parameters must require a reasonable improvement in the predictive ability of the models, otherwise this variable should not be included.

In a first stage of the model fitting, independent equations for the several aboveground biomass components (Ww – wood biomass, Wb – bark biomass, Wbr – branches biomass, Wl – leaves biomass) were fitted using the PROC MODEL procedure of SAS (SAS Institute Inc. 2009a). All the stand/site variables (t, N, S and coppice) were considered to localize the a, b and c parameters and the hypothesis of having them as regressor variables, alone or in combination, in a, b, or c was tested. Since there were a small number of possibilities, all of them where tested, from having a, b and c as linear functions of all the variables to having a, b and c function of only one of the variables. The evaluation of the functions was made considering the SSE (Sum of Squared Errors), MSE (Mean Squared Error) and the Adjusted-R^2 (SAS Institute Inc. 2009b).

Age has been selected as an important predictor but it is a variable that is not accessed as part of current national inventory due to the non-existence of visible annual growth rings in

eucalyptus. Therefore a set of models without this variable, more easy to use in the cases where age is not available, was also developed.

Having data from plots measured in consecutive years, the presence of auto-correlation among the model errors was analysed using the ODS (SAS Institute Inc. 2009c) tool of SAS. The plots of the autocorrelation function (acf) and partial autocorrelation function (pacf) were analysed in order to decide which structure best fitted the data: AR, MA or ARMA and its order.

The heteroscedasticity of the errors was analysed observing the plots of the studentized residuals versus the predicted values. In the case of non-homogeneous variance of the errors, weighted regression was used. The normality of the residuals was assessed with the help of the normal QQ plots and corrected, when necessary, with robust regression using the Huber function (Myers, 1986).

The individual adjustment of the equations does not guarantee that the sum of all the components equals the total aboveground biomass. To ensure compatibility and additivity of the non-linear biomass equations, each biomass component must be estimated by an appropriate allometric equation and the model for total aboveground biomass must be expressed as the sum of the models for all the components, with each one of the components using its specific (Parresol, 1999; Parresol, 2001). The final system of equations was obtained by simultaneous fitting of the equations for each component together with the total aboveground biomass expressed as the sum of all the components. The initial parameters and weight values for the simultaneous fitting were the ones obtained in the individual fitting of each of the equations. The nonlinear seemingly unrelated regressions method described by Bi *et al.* (2010) was used, considering the ITSUR – Iterative Seemingly Unrelated Regressions – option of the PROC MODEL (SAS Institute Inc. 2009c) procedure.

The evaluation of the prediction ability of the models was made using several statistics based on the PRESS residuals (residuals computed by a jack-knife procedure that indicates the predictive ability of the equations by cross validation (Myers, 1986):

- mean value of the PRESS residuals, MPRESS (model bias);

- mean absolute value of the PRESS residuals, MAPRESS (model precision).

The PRESS residuals were computed using the information provided by the ODS tool of SAS. The PRESS residuals were also analysed observing the box-plot graphics of the average and the average of the absolute value of the PRESS residuals as a function of age (t), site index (S) and density (N) classes in order to detect possible tendencies in the bias and/or precision of the equations across these variables.

Results and discussion

Table 2 shows some of the results obtained during the 1st stage of model selection.

Table 2. – Evaluation results of the independent equations for the biomass components (Ww, Wb, Wbr and Wl) using alternative expansions for the b and c parameters.

	parameter	variables		SSE	MSE	Root MSE	R-Square	Adj R-Sq
Ww		without hdom		2087035	929.347	96.403	0.9258	0.9258
		base		44566.6	1.9846	1.4088	0.9984	0.9984
	b	*N*	-	39517.4	1.7598	1.3266	0.9986	0.9986
		coppice	-	41888.8	1.8655	1.3658	0.9985	0.9985
		t	-	38464.6	1.7130	1.3088	0.9986	0.9986
	c	*N*	-	38075.5	1.6956	1.3022	0.9986	0.9986
		coppice	-	41650.9	1.8549	1.3619	0.9985	0.9985
		t	-	38783.3	1.7272	1.3142	0.9986	0.9986
Wb		without hdom		46451.8	20.685	14.382	0.9044	0.9044
		base		9661.6	0.4302	0.6559	0.9801	0.9801
	b	*N*	-	5013.0	0.2232	0.4725	0.9897	0.9897
		coppice	-	7413.3	0.3301	0.5746	0.9847	0.9847
		t	+	9076.4	0.4042	0.6358	0.9813	0.9813
	c	*N*	-	4047.1	0.1802	0.4245	0.9917	0.9917
		coppice	-	7459.5	0.3322	0.5764	0.9846	0.9846
		t	+	9034.9	0.4024	0.6343	0.9814	0.9814
Wbr		without hdom		1588.3	0.0707	0.2659	0.9887	0.9887
		base		1588.0	0.0707	0.2659	0.9887	0.9887
	b	*N*	-	1112.3	0.0495	0.2226	0.9921	0.9921
		coppice	-	1481.9	0.0660	0.2569	0.9895	0.9895
		t	-	1383.1	0.0616	0.2482	0.9902	0.9902
	c	*N*	-	1091.5	0.0486	0.2205	0.9922	0.9922
		coppice	-	1494.0	0.0665	0.2579	0.9894	0.9894
		t	-	1376.2	0.0613	0.2476	0.9902	0.9902
Wl		without hdom		7239.7	0.3224	0.5678	0.9232	0.9232
		base		2703.3	0.1204	0.3470	0.9713	0.9713
	b	*N*	-	2528.6	0.1126	0.3356	0.9732	0.9732
		coppice	-	2701.0	0.1203	0.3468	0.9713	0.9713
		t	-	1682.8	0.0749	0.2738	0.9821	0.9821
	c	*N*	-	2542.0	0.1132	0.3365	0.9730	0.9730
		coppice	+	2701.7	0.1203	0.3469	0.9713	0.9713
		t	-	1761.9	0.0785	0.2801	0.9813	0.9813

Notes:

× the column with signs indicates the signal of the coefficient associated with the variable included in the expansion of the parameter

× the shaded values of SSR indicate that the variable is not significant

× the "base" model refers to the equation formulation $W_i = a\ G^b\ hdom^c$

× the "without hdom" option refers to the equation formulation without hdom, $W_i = a\ G^b$

The inclusion of dominant height as a second regressor in the allometric models with the parameters common for all stands was highly significant for all the biomass components with the exception of branches biomass. Even for this biomass component dominant height was an important regressor when the parameters were localized according to stand characteristics.

Stand density, age and the dummy variable coppice were all important to localize the b and/or c parameters according to the characteristics of the stand (Table 2). The equations selected for each of the biomass components estimation show very good prediction ability with near 99% of the variance explained by the model.

The analysis of the autocorrelation plots showed that the autocorrelation function (acf) rapidly decreased to zero and that the partial autocorrelation function (pacf) abruptly cuts after lag 1, the typical patterns for a first-order autoregressive model - AR (1) for the autocorrelation structure of the errors. The value close to zero of the plotted pacf at lag 1 obtained after adding this autocorrelation structure showed that the AR(1) model removes all the autocorrelation. The AR process expresses the series in terms of past observations and the current disturbances (random error), so it makes sense that this is the error structure that best fits the data. All the equations presented the same behaviour and all of them where adjusted with a AR(1) autocorrelation structure.

The QQ plots showed some evidence of non-normal distribution of the model errors, as it can be seen in Figure 1, so a correction was made using robust estimation with the Huber´s function (Myers, 1986).

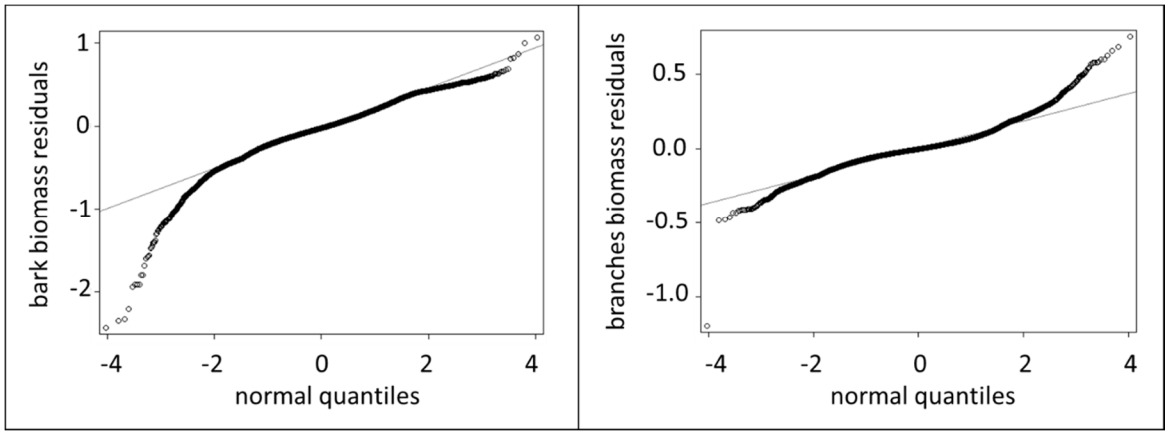


Figure 1. QQ plots for bark (Wb) and branches biomass (Wbr) in Mg ha^{-1}.

Studentized residuals present a close to normal distribution, although with some slightly "heavy" tails, more visible in the case of the wood biomass (Figure 2), but there is no clear violation of the error variance homogeneity assumption.

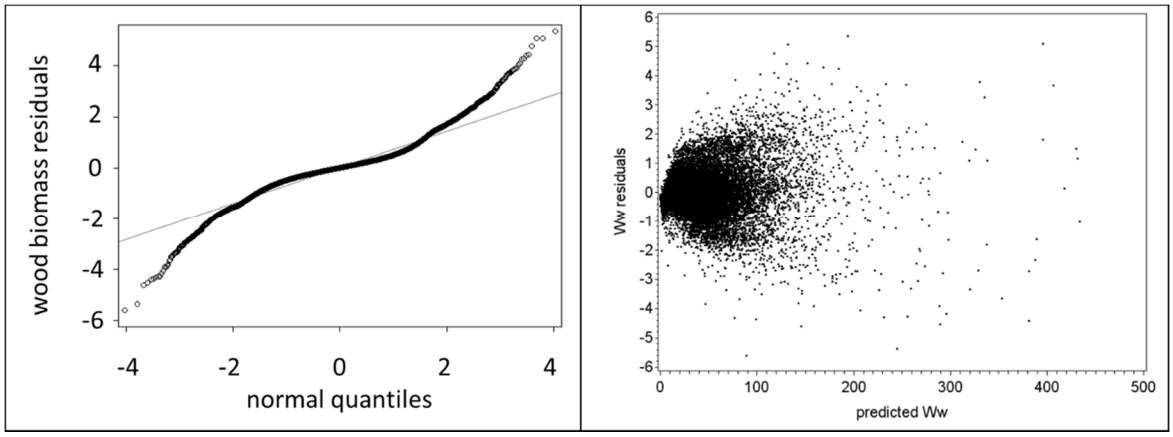


Figure 2. QQ plot and graphic of the studentized residuals over predicted values of wood biomass (Ww) in Mg ha^{-1}.

Table 3. shows the final equations and the respective parameters estimates that resulted from the simultaneous adjustment of the two systems of equations, with or without one of the parameters localized according to stand age, and Table 4. shows the values of the fitting and predictive ability statistics used to evaluate the models. Both systems of equations present high values of Adjusted-R^2 for all the component equations.

Table 3 – Estimated parameters for the component biomass equations of the two systems developed, with or without one of the parameters localized according to stand age. (continues)

$\mathrm{Ww} = \mathrm{a}\,\mathrm{G}^{\,b}\,\mathrm{hdom}^{\,c}$	$\mathrm{a} = 0.101294$	With age
	$\mathrm{b} = 1.033164 - 0.97731\frac{t}{1\,000} - 0.45457\frac{N}{100\,000} - 0.00212\,coppice$	
	$\mathrm{c} = 1.179353$	
	$\mathrm{a} = 0.106977$	Without age
	$\mathrm{b} = 1.021305$	
	$\mathrm{c} = 1.166699 - 0.7366\frac{N}{100\,000} - 0.00111\,coppice$	

Table 3 – Estimated parameters for the component biomass equations of the two systems developed, with or without one of the parameters localized according to stand age. (continued)

$Wb = a\,G^{\,b}\ hdom^{\,c}$	$a = 0.037758$ $b = 1.163814 - 0.01275\ coppice + 3.154601\frac{t}{1\,000}$ $c = 0.720653 - 0.45926\frac{N}{10\,000}$	With age
	$a = 0.034115$ $b = 1.20028 - 0.01071\ coppice$ $c = 0.758052 - 0.54483\frac{N}{10\,000}$	Without age
$Wbr = a\,G^{\,b}\ hdom^{\,c}$	$a = 0.413399$ $b = 1.010733 - 0.00263\ coppice$ $c = 0.04027 - 0.15766\frac{N}{10\,000} - 4.15935\frac{t}{1\,000}$	With age
	$a = 0.488369$ $b = 0.997656$ $c = -0.04404 - 0.15047\frac{N}{10\,000} - 0.00241 coppice$	Without age
$Wl = a\,G^{\,b}\ hdom^{\,c}$	$a = 1.709464$ $b = 1.054603 - 0.9485\frac{N}{100\,000} - 5.88885\frac{t}{1\,000}$ $c = -0.47726$	With age
	$a = 2.056249$ $b = 0.987647 - 0.804\frac{N}{100\,000}$ $c = -0.53111$	Without age

Table 4 – Fitting and predictive ability statistics for the two systems of equations, with or without one of the parameters localized according to stand age, fitted through nonlinear ITSUR

	SSE	MSE	Root MSE	R^2_{adj}	MPRESS	MAPRESS	age
Ww	26658.0	1.1871	1.0896	0.9991	-0.0197	0.53012	yes
	26893.3	1.1976	1.0943	0.9990	0.01131	0.53568	no
Wb	4641.9	0.2067	0.4547	0.9904	0.00144	0.05472	yes
	6543.7	0.2914	0.5398	0.9865	0.02048	0.07934	no
Wbr	448.5	0.0200	0.1413	0.9968	0.05059	0.39133	yes
	654.9	0.0292	0.1708	0.9953	0.06759	0.39442	no
Wl	973.8	0.0434	0.2082	0.9897	-0.0026	0.05983	yes
	1400.3	0.0624	0.2497	0.9851	0.00525	0.09356	no

SSE: Sum of squared errors; MSE: Mean squared error; R^2_{adj} – adjusted R-squared; MPRESS: mean value of PRESS residuals; MAPRESS: mean absolute value of PRESS residuals

The plots of the observed versus predicted values (Figure 3) show a close to linear relationship, closely following the 1:1 line, with some dispersion, more evident in the case of the leaf biomass equations. In the system of equations without age it is visible that the predictions are not so accurate, especially in the case of branches and bark biomass.

The analysis of the PRESS residuals over stand density, site index and age classes shows no relevant bias or tendencies.

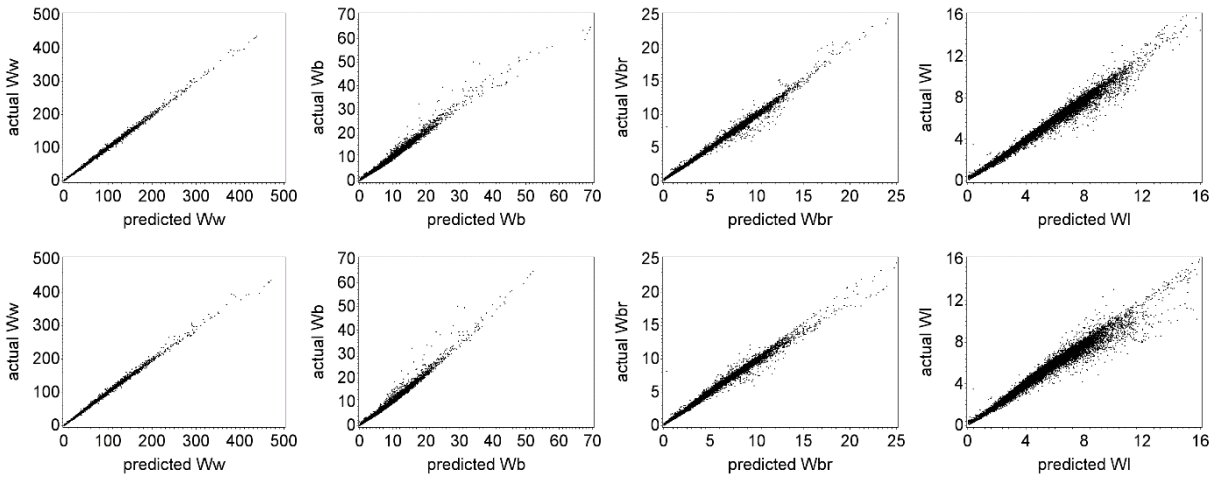


Figure 3. Plots of observed versus predicted values for the four variables modelled (Ww, Wb, Wbr and Wl (Mg ha^{-1}) from left to right) for the models with age (above) and without this variable (below).

Discussion

The objective of the present research was to develop tools to predict total aboveground biomass and biomass per tree component at stand level using as regressors variables that are usually available from forest inventories. Biomass estimates from field plots are usually estimated by summing up tree biomass estimates obtained with individual tree allometric equations. Stand level prediction equations are not common in the literature, but a study undertaken in Australia (Snowdon *et al.*, 2000) showed evidence that application of 'stand-scale' equations is likely to be as effective as the use of more site- and species-specific equations applied to individual trees in a stand.

Stand basal area is a very important variable for predicting biomass (Snowdon, 1992; Husch *et al.*, 2003; Bi *et al.*, 2010) and it has been used often along with dominant height to predict volume, which is closely related to wood biomass. The proposed models were based on an allometric relationship between each biomass component and stand basal area and dominant height. González-García *et al.* (2013) obtained very good models for *Eucalyptus nitens* with a simple allometric model that uses just basal area as regressor. Models also including dominant height lead to a slight improvement of model accuracy. In our case dominant height really improved biomass prediction for all the biomass components. The large range of stand ages (0.6-33.9) and stand densities (67-8563) observed in our data set may be a justification for the need to include dominant height as a second regressor. António *et al.* (2007) also found, when developing tree level biomass equations for *E.globulus*, that tree height or tree crown length (for stem and crown biomass, respectively) lead to more precise biomass estimations than the use of the simple dbh based allometric model. Additionally our study showed that there is a significant advantage in localizing the model parameters of the allometric model by expressing them as a linear function of the characteristics of the stand, namely stand density and age. The use of different parameters for seedling and coppice stands, taken into account through a dummy variable for coppice, was also significant.

Age has an important effect on some biomass components, such as bark and foliage, because of the importance they have on younger stands and the small importance in terms of biomass components in older stands (António *et al.*, 2007; Bi *et al.*, 2010). A somehow poorer performance in the system of equations that do not localize the parameters according to stand age was expected and our results proved this hypothesis.

Stand density is another stand characteristic with a strong impact on biomass and this fact is also patent in the significant reduction of the SSE when the parameters are localized as a function of stand density.

Conclusion

Two systems of equations to predict stand level aboveground biomass and biomass per tree component for even-aged stands of planted and coppiced *Eucalyptus globulus* were developed. The equations are simple and based on a small number of variables readily available from operational forest inventory, which makes them easy to apply and with a large use potential. The equations are precise and with a high predictive ability.

The development of the systems of equations that are the main result of this paper lead also to the following conclusions:

1. Total aboveground biomass and biomass per tree component are well estimated using an allometric relationship with stand basal area and dominant height but the parameters must be localized according to the characteristics of each particular stand
2. There is the need to use different allometric constants for planted and coppiced stands

Acknowledgments

This paper is a part of the PhD project of the first author, which is funded by a scholarship (SFRH/ BD/ 46594/ 2008) granted by Fundação para a Ciência e Tecnologia (Portugal).

References

António, N., Tomé, M., Tomé, J., Soares, P., Fontes, L. (2007). Effect of tree, stand and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research 37: 895-906.

Attiwill, P.M., Ovington,J.D. (1968). Determination of forest biomass. Forest Science 14: 13-15.

Barrio-Anta, M., Balboa, M., Castedo, F., Diéguez-Aranda, U., Álvarez-González, J.G. (2006). An ecoregional model for estimating volume, biomass and carbon pools in maritime pine stands in Galicia (northwestern Spain). Forest Ecology and Management 223(1-3): 24-34.

Baskerville, G.L. (1965). Estimation of dry weight of tree components and total standing crop in conifer stands. Ecology 46: 867-869.

Bettinger, P., Hayashi, R. (2006). Estimation of above-ground biomass with remotely sensed imagery: A brief literature review. Research Note 25, Center for Forest Business, University of Georgia.

Bi, H., Long, Y., Turner, J., Lei, Y., Snowdon, P., Li, Y., Harper, R., Zerihun, A., Ximenes, F. (2010) Additive prediction of aboveground biomass for *Pinus radiata* (D. Don) plantations. Forest Ecology and Management 12:2301– 2314

Brown, S.L., Schroeder, P., Kern, J.S. (1999). Spatial distribution of biomass in forests of the eastern USA. Forest Ecology and Management 123: 81–90.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., Bailey, R.L. (1983). Timber management: a quantitative approach. Florida: Krieger Publishing Company.

Crow, T.R. (1971). Estimation of biomass in an even-aged: regression and "mean tree" techniques. In: Young, H.E. [ed.], Forest biomass studies. IUFRO Yield and Growth Working Group on forest biomass studies. 25th IUFRO Congress: 35-47.

Faias, S.P., Paulo, J., Soares, P., Patrício, M.S., Freire, J.P., Tomé, M. (2009). Modelling biomass expansion factors for the most important forest tree species in Portugal. Dissertation, Universidade Técnica de Lisboa; p. 18–38.

Fang, J.Y., Chen, A.P., Peng, C.H., Zhao, S.Q., Ci, L.J. (2001). Changes in forest biomass carbon storage in China between 1949 and 1998. Science 292: 2320–2322.

González-García, M., Hevia, A., Majada, J., Barrio-Anta, M. (2013). Above-ground biomass estimation at tree and stand level for short rotation plantations of *Eucalyptus nitens* (Deane & Maiden) Maiden in Northwest Spain. Biomass and Bioenergy 54: 147-157.

Gorbatenko, V.M. (1970). Biological productivity of the pine phytocoenoses related to climatic conditions in their growth regions. PhD Dissertation. Institute of Forest and Wood. Siberian Branch. USSR Academy of Sciences. Krasnoyarsk. (In Russian.)

Husch, B., Beers, T.W., Kershaw, J.A. (2003) Forest mensuration, 4th edn. Wiley, New York

Kuzikov, I.E. (1979). Changing phytomass in the green moss fir forests of in different climatic conditions of Middle Siberia. PhD Dissertation. Institute for Forest Research. Krasnoyarsk. (in Russian).

Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing. 27: 1297-1328.

Madgwick, H.A.I., Satoo, T. (1975). On estimating the aboveground weights of tree stands. Ecology 56: 1446-1450.

Monserud, R.A., Onuchin, A.A., Tchebakova, N.M. (1996) Needle, crown, stem, and root phytomass of P. sylvestris stands in Russia. Forest Ecology and Management 82:59–67

Myers, R. H. (1986). Classical and Modern Regression with Applications. Second edition. Duxbury Classic Series. 357 pp.

Oliveira, T., Tomé, M. (2015). A system of compatible models to predict and project crown ratio and crown length for eucalypt plantations. Submitted.

Parresol, B.R. (1999) Assessing tree and stand biomass: a review with examples and critical comparisons. Forest Science 45:573–593

Parresol, B.R. (2001) Additivity of nonlinear biomass equations. Can J For Res 31:865–878

SAS Institute Inc. (2009a) SAS ® 9.2 Output Delivery System: User's Guide. Cary, NC: SAS Institute Inc. 984 p.

SAS Institute Inc. (2009b) SAS/ETS ® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 2876 p.

SAS Institute Inc. (2009c) SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc. 7886 p.

Schroeder, P., Brown, S., Mo, J., Birdsey, R., Cieszewski, C. (1997). Biomass estimation for temperate broadleaf forests of the United States using inventory data. Forest Science 43: 424-434.

Semechkina, M.G. (1978). Structure of the Scotch pine stand's phytomass. Novosibirsk: Nauka. (in Russian).

Snowdon, P. (1992) Ratio methods for estimating forest biomass. N Z J For Sci 22:54–62

Snowdon, P., Eamus, D., Gibbons, P., Khanna, P., Keith, H., Raison, J., Kirschbaum, M. (2000). Synthesis of allometrics, review of root biomass and design of future woody biomass sampling strategies. In: The Australian Greenhouse Office, Canberra, National Carbon Accounting System Technical Report 17:22-29.

Soares, P. Tomé, M. (2001). A tree crown ratio prediction equation for eucalypt plantations. Annals of Forest Science 58(2): 193-202.

Soares, P., Tomé, M. (2012) Biomass expansion factors for *Eucalyptus globulus* stands in Portugal. Forest Systems, 21 (1), pp. 141–152

Teobaldelli, M., Somogyi, Z., Migliavacca, M., Usoltsev, V. (2009). Generalized functions of biomass expansion factors for conifers and broadleaved by stand age, growing stock and site index. Forest Ecology and Management, 257 pp. 1004–1013

Tomé, M., Oliveira, T.S., Soares, P. (2006). O modelo GLOBULUS 3.0. Dados e equações. Publicações GIMREF - RC2/2006. Centro de Estudos Florestais, Instituto Superior de Agronomia, Lisboa.

Tomé, M., Ribeiro, F., Faias, S. (2007). Relação hipsométrica geral para *Eucalyptus globulus* Labill. em Portugal. Silva Lusitana 15(1): 41-56.

Tomé, M., Ribeiro, F., Soares, P. (2001). O modelo GLOBULUS 2.1. Relatórios Técnico-Científicos do GIMREF nº 1/2001. Centro de Estudos Florestais, Instituto Superior de Agronomia, Lisboa.

Turnbull, C.R.A., McLeod, D.E., Beadle, C.L., Ratkowsky, D.A., Mummery, D.C., Bird, T. (1993). Comparative early growth of Eucalyptus species of the subgenera Monocalyptus and Symphyomyrtus in intensively-managed plantations in southern Tasmania. Australian Forestry 56(3): 276-286.

Usol'tsev, V.A. (1988). Growth and Phytomass Structure of Stands. Novosibirsk: Nauka. (in Russian).

Zianis, D., Muukkonen, P., Mäkipää, R. & Mencuccini, M. (2005) Biomass and stem volume equations for tree species in Europe. Silva Fennica Monographs, 4, 1–63.

Chapter 3 - Modelling the diameter distribution in planted and coppiced *Eucalyptus globulus* plantations

Executive summary

A general model for diameter distribution of eucalyptus stands in Portugal was fitted and equations for average, median, minimum and maximum diameters were developed.

Abstract

- Context: The diameter distribution is one of the most important methods of describing stand structure. All the other tree variables, like height, volume or biomass are strongly correlated with diameter. Diameter distribution modelling is therefore important as a complement to whole stand models as a contribution to indicate the potential use of forest stands by allowing the prediction of assortments or as support to an efficient harvest planning. Diameter distribution modelling is also needed for the initialization of individual tree models.

- Aims: Diameter distribution modelling is usually made by using probability density functions (pdf´s) for which parameters need to be estimated. The diameter distribution modelling may be achieved by modelling each parameter of the pdf as a function of stand variables (parameter prediction method) or by using the method of moments to recover the parameters of the pdf (parameter recovery method). The research presented here presents the modelling of the diameter distribution in planted and coppiced *Eucalyptus globulus* plantations with the Johnson-SB pdf and the methods of moments.

- Methods: The Johnson-SB probability density function and the method of moments was used in this work. Equations for average, median, minimum and maximum diameters were developed. The equations ensure compatibility among the different moments and were simultaneously fitted as a nonlinear seemingly unrelated regression system using the ITSUR option of the SAS PROC MODEL procedure

- Results: The different diameters can be estimated with simple and generally available stand level variables: stand density, basal area, dominant height, quadratic mean diameter and a dummy variable that indicates if the stand is high-forest or coppice.

- Conclusion: Using the equations developed in this work together with predictions of basal area, dominant height, stand density and quadratic mean diameter allows the use of a program implemented in the sIMFLOR interface that estimates the parameters of a Johnson-SB pdf. With this parameter values, the diameter distribution of any plot can be simulated.

Keywords: diameter distribution, Johnson-SB probability density function, method of moments, parameter recovery

Introduction

The diameter distribution is one of the most important methods of describing stand structure. All the other tree variables, like height, volume or biomass are strongly correlated with diameter. Diameter distribution modelling is therefore important as a complement to whole stand models as a contribution to indicate the potential use of forest stands by allowing the prediction of assortments or as support to an efficient harvest planning. Diameter distribution modelling is also needed for the initialization of individual tree models.

Diameter distributions are usually modelled using probability density functions such as the Weibull, the truncated Weibull, the beta or the Johnson SB (e.g. Hafley and Schreuder, 1977; Strub and Burkhart, 1975; Hyink, 1980; Hyink and Moser, 1983; Palahi *et al.*, 2007; Fonseca *et al.*, 2009; Mateus and Tomé, 2011). Two steps must be considered when modelling diameter distributions: 1) selection of a probability density function (pdf) and 2) modelling the diameter distribution with the selected pdf. The first step is usually done by analysing the empirical skewness ($\sqrt{\beta 1}$) and kurtosis ($\beta 2$) observed in a data set covering all the structures relevant for the species in the $\beta 1$-$\beta 2$ space. The different pdfs are represented in this space and can be compared with the observed values. The diameter distribution modelling may be achieved by modelling each parameter of the pdf as a function of stand variables (parameter prediction method) or by using the method of moments to recover the parameters of the pdf (see, for example, Burkhart and Tomé (2012) for more details on both the selection of a pdf and the development of diameter distribution models).

Diameter distribution models have been mainly associated with empirical models, although they can be added to any type of stand model. Landsberg *et al.* (2005) analysed the appropriateness of the Weibull distribution to model stem size distributions in *Pinus sylvestris* stands in Finland and, from the 3-PG outputs, selected the mean stem diameter as predictor of the Weibull parameters using polynomial or allometric models, but with limited success (an R^2 value of 0.96 was obtained for the location parameter but smaller values of R^2 – 0.34 and 0.07 – were obtained for the shape parameters).

In a study undertaken for planted *Eucalyptus globulus* Portugal, the Johnson's distribution has been shown as the most appropriate (Furtado, 1997). Mateus and Tomé (2011) used this distribution and the parameter recovery method to model the diameter distribution of first-rotation eucalyptus plantations in Portugal.

The objective of the research described here was to develop a general model for diameter distributions of *Eucalyptus globulus* stands in Portugal covering the whole area of distribution of the species in Portugal and including planted and coppiced stands.

Methods

II.1 Data

Data for this study came from the large database available in Portugal for eucalyptus stands. It includes data from permanent plots, continuous forest inventory of the pulp companies and experimental trials and covers almost all the regions with a significant area of eucalyptus stands in Portugal. It contains data from planted and coppiced stands, part of it consisting of successive measurements (usually with 1 year interval between them) on the same plots which means that the presence of auto-correlation has to be considered. On each plot every tree had its diameter at breast height (d) measured. In current forest inventories, usually only trees with d bigger than 5 cm are measured, but on permanent plots and trials all trees are measured, which means that there is information on very small trees (Table 1).

Table 1 – Characterization of the data used in the diameter distribution modelling (number of plots=13 867; n=22 364)

	Minimum	Mean	Maximum
Minimum diameter	0.1	4.04	19.15
Median diameter	0.45	11.13	27.5
Average diameter	0.56	10.77	26.10
Maximum diameter	1.9	17.85	41.5

Tree level information was considered and for each plot the minimum (d_{min}), median (d_{med}), average ($\bar{d}$) and maximum (d_{max}) diameters where calculated, as well as other stand level variables. After analysis and validation of the values, the data set used contains information on more than 22 000 observations taken in 13 867 plots covering a wide range of ages, stand densities and site index (Table 2).

Table 2 – Distribution of the data by age, site index and stand density

S, base age 10 (m)	Stand density (trees/ha)	Age				Total
		t ≤ 4	4 < t ≤ 8	8 < t ≤ 12	t > 12	
S ≤ 16	N < 1111	180	1000	1 631	138	2 949
	1111 ≤ N < 1667	54	435	679	62	1 230
	N > 1667	10	68	132	106	316
16 < S ≤ 20	N < 1111	314	1 966	2 582	283	5 145
	1111 ≤ N < 1667	278	1 488	1 602	139	3 507
	N > 1667	98	283	480	78	939
20 < S ≤ 24	N < 1111	156	1 171	1 568	229	3 124
	1111 ≤ N < 1667	199	948	1 274	155	2 567
	N > 1667	127	270	416	89	902
S ≥ 24	N < 1111	36	186	387	81	690
	1111 ≤ N < 1667	55	199	367	75	696
	N > 1667	33	84	128	45	290
Total		1 540	8 098	11 246	1 480	**22 364**

II.2 Model selection and fitting

The diameter distribution modelling used the parameter recovery method that has been shown to have better performance than the parameter prediction method (see e.g. Burkhart and Tomé 2012). This method, based on the method of moments (see Mateus and Tomé, 2011 for details) requires information on a set of moments whose prediction requires the following stand variables: stand basal area (G), number of trees per ha (N), average, median, minimum and maximum diameters ($\bar{d}$, d_{median}, d_{min}, d_{max}). To ensure that the estimates respected well-established principles, these authors imposed some restrictions in the construction of the models: minimum diameter should be lower than all the others and greater than zero and the average diameter should be lower than the quadratic mean diameter. The minimum and average diameters were therefore modelled as:

(1) $d_{min} = \bar{d} - (\text{proportion of } \bar{d})$, (alternatively d_{median} can be used instead of $\bar{d}$)

(2) $\bar{d} = d_g - (\text{proportion of } d_g)$, with d_g – quadratic mean diameter

In both equations the proportions depend on the characteristics of the stand.

The median and maximum diameters were modelled with multiple allometric models, allowing the allometric constant or the scale parameters to be expressed as a function of stand age (t), stand density (N), dominant height (hdom) or cutting cycle.
The system of equations used was:

(3) $$d_{min} = \min(\bar{d}, d_{median}) - \left(\frac{1}{1+\exp(\sum a_i X_i)} \right) \min(\bar{d}, d_{median})$$

(4) $$d_{median} = k \prod X_i^{a_i}$$

(5) $$\bar{d} = d_g - \left(d_g \frac{1}{1+\exp(\sum a_i X_i)} \right)$$

(6) $$d_{max} = k \prod X_i^{a_i}$$

where X_i represents stand variables such as G, hdom, $\bar{d}$ or d_g and the scale parameter k and/or the allometric constants a_i can be expressed as a function of the remaining the variables.
The present work used the same system of equations, but testing different combinations of stand variables. Site index (S) was used as regressor in the equations develop by Mateus and Tomé (2011), but this variable presents some problems. S is a measure of the potential growth at a site and is usually expressed as dominant height at age 10 in eucalyptus stands. Implicit in this concept are the assumptions that the dominant height growth of a stand is independent of changes in the environment and is not influenced by stand stocking. When growth conditions are stable, one can use this concept without problems, but when the interest falls on new locations, changing soil or climatic conditions, the growth response is highly non-linear and the relationships can be very difficult to predict. Differences in soil or climatic conditions will change the development rate of trees and consequently the value of the dominant height. Also, differences in silvicultural techniques will also have impact on the tree growth and may change the dominant height curve. The site index concept is therefore strongly connected to the site specific conditions of the data used which presents problems not only when there are no previous data available for that site, but also when one wants to consider climate changes, for example. Due to all these problems, it was decided not to use site index as regressor.
Stand age is difficult to obtain in forest inventories as the growth rings are not well defined, therefore it is only available for eucalyptus stands where there is information about plantation dates. If that information is not available, it is very difficult, or even impossible, to estimate it by an analysis of the trees and stand. Therefore the impact of having stand age as regressor variable was always carefully considered and compared to the alternative of not having this variable in the equations.
The candidate regressors included a dummy variable for coppice (1 for seedling stands; 2 for the ones regenerated by coppice) and in most equations, a scaling factor (N/1000 or

coppice/10, for example) was used so that the coefficient estimates where, whenever possible, on the same scale (Schabenberger and Pierce, 2002).

On a first stage, a visual analysis was conducted on the plots of each of the diameters (d_{min}, $\bar{d}$, d_{med} and d_{max}) versus the different variables to be considered (dg, t, N, G, hdom and coppice). We are aware of the multi-correlations that exist among the different stand variables tested as possible regressors and therefore this analysis was considered as just exploratory of the variables with more relationship to the diameters that are the object of modelling. Then, using the PROC NLIN procedure (SAS Institute Inc., 2009b) the parameters of the equations were expressed as a linear function of stand variables, starting with each of the variables alone and going to different combinations of stand variables to see the impact of each one, alone or in combination, in the value of the residual sum of squares:

(7) $\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)$,

where y_i – the i^{th} value of the variable to be predicted; $\hat{y}_i$ – predicted value of y_i.

Considering the results of this first stage, the best models where then adjusted using the PROC MODEL (SAS Institute Inc., 2009a) procedure and compared based on several statistics of fit and prediction ability. The fitting was assessed by the residual sum of squares (SSE), the mean square error (MSE) and the Adjusted-R^2 (SAS Institute Inc., 2009b).

The models were analysed considering the regressor variables, the impact on the fit statics of their presence or absence in the equations and considering also the plots previously made. The objective was to keep the equations as simple as possible, based on regressor variables easy to obtain while maintaining a good prediction ability.

The heteroscedasticity of the errors was analysed observing the plots of the studentized residuals versus the predicted values. In case of nonhomogeneous variance of the errors, weighted regression was used. The residuals were weighted, in an iterative process, with the inverse of the standard error of the estimate. The normality of the residuals was analysed with the help of the normal QQ plots and corrected, when necessary, with robust regression using the Huber function (Myers, 1986).

The presence of auto-correlation among the model errors (due to the existence of repeated measurements on the same plot) was analysed using the ODS tool of SAS (SAS Institute Inc., 2009c). The plots of the autocorrelation function (acf) and partial autocorrelation function (pacf) were analysed in order to decide which structure best fitted the data: AR, MA or ARMA and its order.

After the individual adjustments, the equations were simultaneously fitted to ensure the model compatibility and because the independent variables appear in more than one functions. Since the set of equations has contemporaneous cross-equation error correlation (also known as

nonlinear seemingly unrelated regression system), the fitting was made using PROC MODEL (SAS Institute Inc., 2009a) and the ITSUR option, which provides an iterative estimation of the cross-equation covariance matrix. The initial parameters and weight values for the simultaneous fitting were the ones obtained in the individual fitting of each of the equations.

II.3 Model evaluation

Model evaluation begins in the theoretical aspects of model building and the logic of its structure and of the biological aspects. Besides from that, models should be validated in other ways and visual/graphical inspection of the model predictions and of the residuals is one of the most effective ways of doing it. Since the model structure was already considered in the development of the equations, just the logic of the variables selected of regressors was analysed. To evaluate the prediction ability, several statistics based on the PRESS residuals (residuals computed by a jack-knife procedure that indicate the predictive ability of the equations by cross validation (Myers, 1986)) where considered:

1. mean value of the PRESS residuals, MPRESS (model bias);
2. mean absolute value of the PRESS residuals MAPRESS (model precision).

Box-plot graphics of the PRESS residuals were also made considering classes of t, S and N and the average and average of the absolute value of the PRESS residuals were both calculated to detect possible tendencies in the bias and/or precision of the equations. To further validate the equations, observed values where plotted against the correspondent estimated values and the bias on the observed-to-predicted fit was tested examining if the values where or not very distant from the 1:1 reference line of the graphic.

Results

As mentioned in the methods section, stand basal area (G), dominant height (hdom) and quadratic mean diameter (dg), variables that are needed to apply the parameter recovery method, were not modelled in this research as these variables are usually part of the output of whole stand models. The results report the modelling of the remaining diameter variables needed: minimum, average, median and maximum diameters (d_{min}, $\bar{d}$, d_{med} and d_{max}). The first analysis of the plots of the different diameters versus the candidate regressor variables (Figures 1 to 4) showed that there are evident relationships between the diameters and dg, an expected result since dg is the quadratic mean diameter. All the other diameters asides from d_{min} ($\bar{d}$, d_{med} and d_{max}) have some relationship with hdom, G and t. Minimum diameter is the

variable with less evident relationships with other stand variable, with the exception of the relationship with $\bar{d}$ and d_{med} that are approximately linear.

III.1 Minimum diameter

The equation that resulted from the work of Mateus and Tomé (2011) considered that d_{min} must be modelled by considering the minimum of ($\bar{d}$, d_{med}) decreased by a certain value (a percentage of the variable used as the basis). That percentage that is subtracted is a function of the different variables considered and it ensures that the minimum diameter is in fact smaller than the other diameters. In the present research we preferred to test both variables instead of basing the model in the minimum of ($\bar{d}$, d_{med}).

The first analysis of the plots of $\bar{d}$ and d_{med} versus d_{min} (Figure 1) showed a similar relationship of d_{min} with both variables, but the results from the tests made with the PROC NLIN procedure show (Table A.1. in the appendices) that d_{min} relates better to $\bar{d}$ than to d_{med} and that using $\bar{d}$ as a the basis variable gives better predictions. The value of SSE for the full model using d_{med} is 30376.9, while the value for the full model using $\bar{d}$ is 27041.6 and the respective values of R^2 are 36% versus 28%, so no further tests where made using d_{med} as the basis variable.

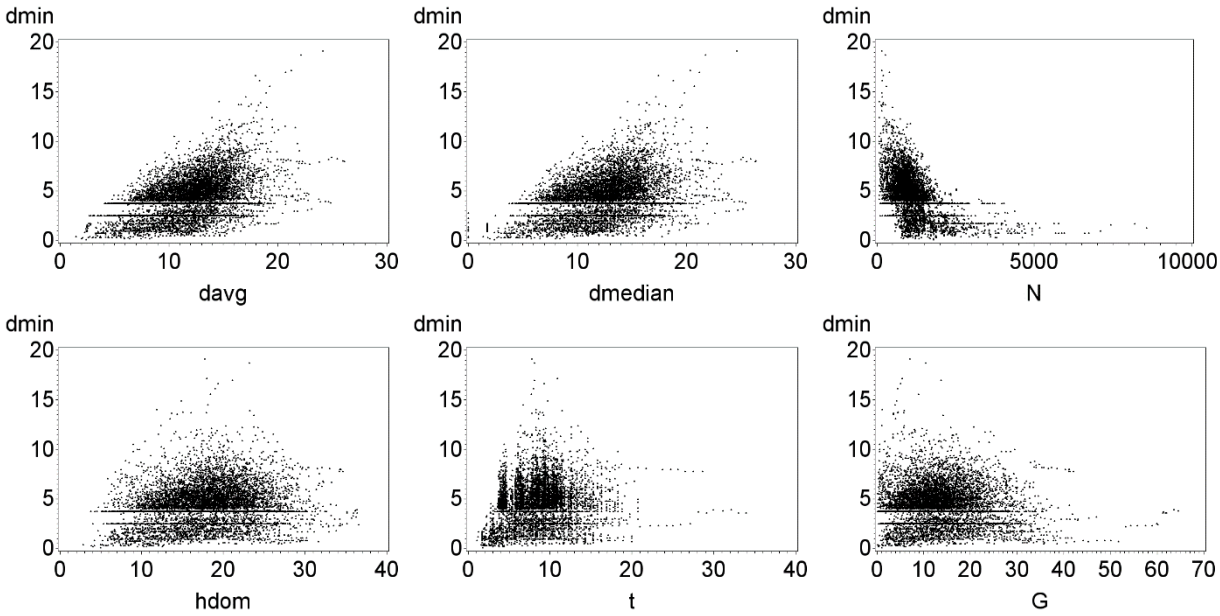


Figure 1. Plots of minimum diameter (cm) versus the different candidate regressor variables: average diameter (cm), median diameter (cm), stand density (ha^{-1}), dominant height (m), age (years) and basal area (m^2 ha^{-1}).

The variables that individually reduce the value of the SSE the most are G and hdom, and the one that shows a smaller impact when considered alone is N. When considering two variables together, G and N are two of the best variables, while the parameter associated with hdom sometimes becomes non significantly different from zero.

The equations that presented the lowest values of SSE, the ones that had as variables (N, hdom, coppice, G) and (N, coppice, t, G) were adjusted with the PROC MODEL procedure and considering the autocorrelation structure and weighted errors. On the first equation, the parameter associated with hdom was not significantly different from zero. A similar result was obtained for the second equation but with the variable t, showing that the model with just 3 variables might be just as good. The model with (N, hdom and coppice) was then tested and the value of the SSE was very similar to the ones obtained with the previous models (27148.0 versus 27135.2 and 27041.6 respectively for the models with 4 parameters), so the final equation selected was:

(8) $$d_{min} = \bar{d} - \left(\bar{d} \ \frac{1}{1+\exp(a1+ a2\,N + \ a3\,coppice + a4\,G)}\right)$$

III.2 Average diameter

The average diameter equation has a structure similar to the one of the minimum diameter and the idea behind it is to start with the quadratic mean diameter and then subtract from it a part of that diameter that is a function of other stand variables. Looking at the plots with the different candidate regressor variables (Figure 2), $\bar{d}$ shows a good and almost linear relation with dg, which confirms the choice for this variable as the main variable of the function. From the other variables, hdom is the one that shows a better relationship with $\bar{d}$, although t and G also present a good relationship.

In terms of the models fitted in the first stage of model selection (Table A.1. in the appendices), the variable that implies a larger reduction in the SSE is N, the variable whose relationship with the average diameter was less evident in the plots. This result is explained by the correlation that exists between the several candidate regressor variables and shows the importance of the second stage used for model development. Considering the variables all together, in the full model the parameter associated with hdom is not significantly different from zero, therefore simpler models were analysed. The model without hdom and with (coppice, t and G) is the one that presents the best results in terms of SSE value, 2986 the same value as the full model. As mentioned before, stand age is not always an easy variable to access, so the model without this variable was also considered, always keeping in mind that there is a relationship between

this variable and $\bar{d}$ and expecting that not using it leads to higher SSE and will tend to lower the predictive ability of the model, although the others variables present in the equation will tend to compensate this absence. In reality, the model without t has a 0.9845 value of adjusted R-square, versus the 0.9856 value of the model with t.

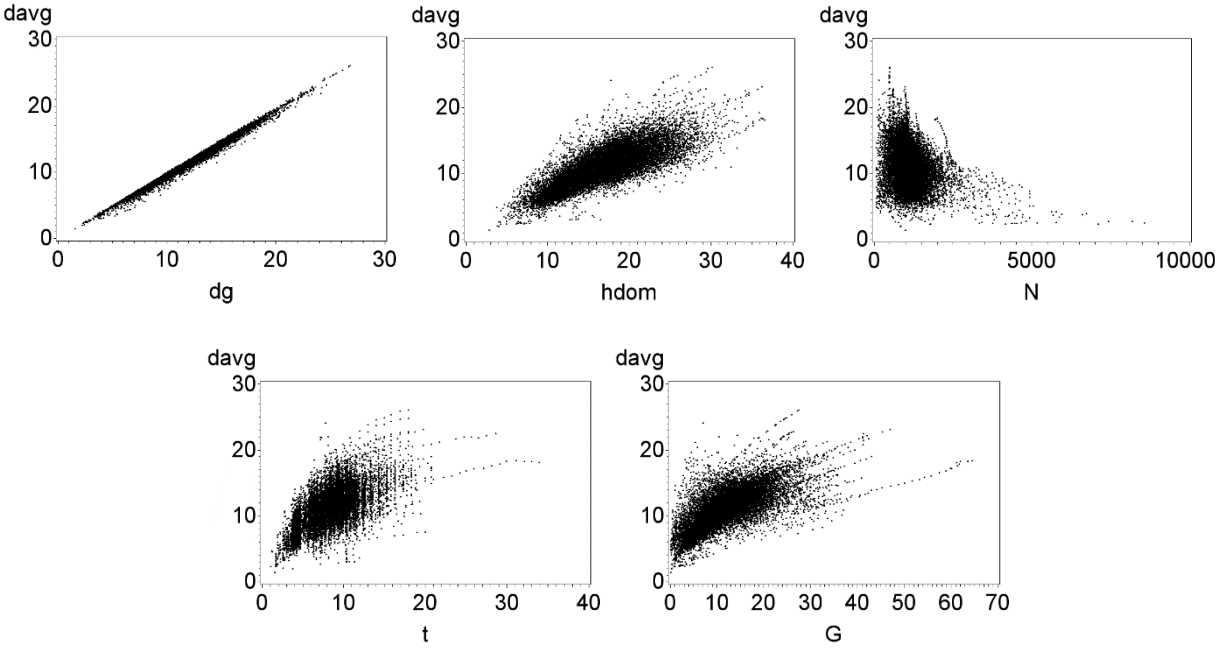


Figure 2. Plots of average diameter (cm) versus the different candidate regressor variables: quadratic mean diameter (cm), stand density (ha^{-1}), dominant height (m), age (years) and basal area (m^2 ha^{-1}).

The QQ-plots, the plots of residuals over predicted values and the plots of actual versus predicted values are so similar in both models that it was decided that there was no significant advantage on using t in the prediction equation. The final equation considered was then:

$$\bar{d} = dg - \left(dg \frac{1}{1+\exp\left(a1 + a2\frac{N}{1000} + a3\frac{G}{100} + a4\frac{coppice}{10}\right)} \right) \quad (9)$$

III.3 Medium diameter

Observing the plots of d_{med} versus the different variables (Figure 3) it can be seen that there is a clear and strong linear relationship between d_{med} and dg. There is also an evident relationship between d_{med} and hdom, which is reflected in the analysis undertaken in the first stage of model development (Table A.2. in the appendices).

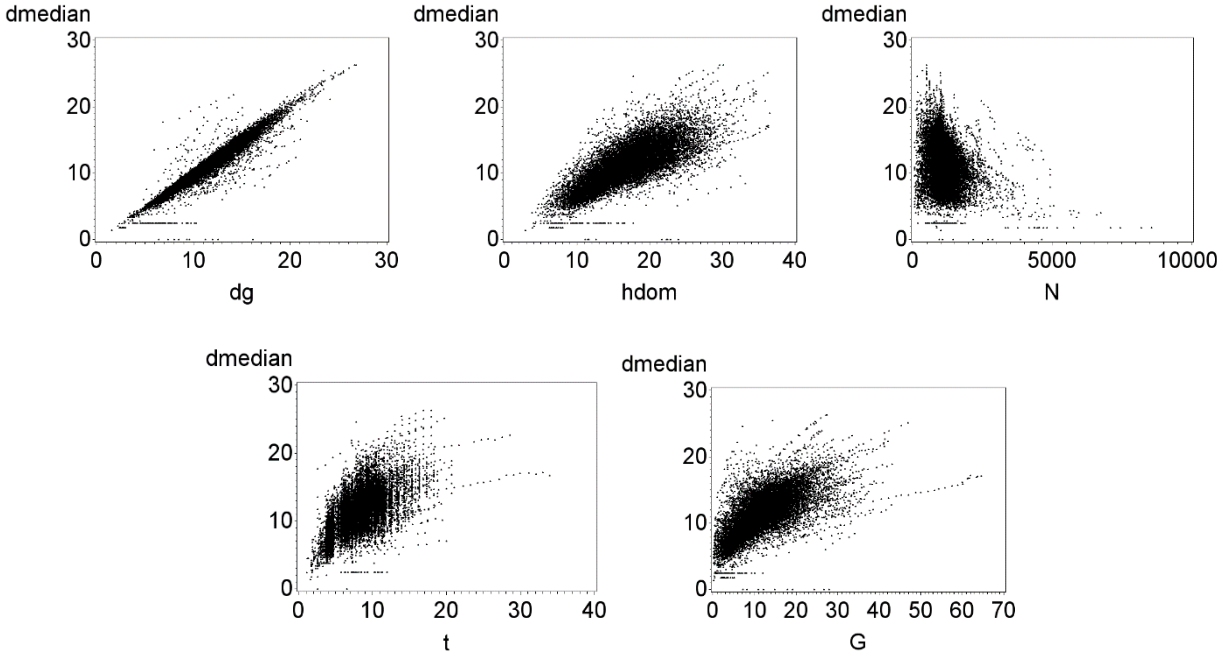


Figure 3. Plots of median diameter versus the different variables: quadratic mean diameter (cm), stand density (ha^{-1}), dominant height (m), age (years) and basal area ($m^2\ ha^{-1}$).

The model that uses only dg (93% modelling efficiency for the full model) performs better than the one using only hdom (80% modelling efficiency for the full model) and there is some significant improvement in using hdom and dg together. There is also an evident relationship between d_{med} and t and G and the models using only these variables are the ones that present the lower values of SSE. In the plots N doesn´t have such a clear relationship, but it is present in the models with the best results. The parameter associated with the coppice variable does not significantly differ from zero in most of the equations, so the selected equation uses as predictor variables (dg, hdom, G, N and t) and has a modelling efficiency value of 0.926 and a SSE value of 17497. Age is once more a variable that is important for the models and once more the presence or absence of this variable was analysed. In the model without this variable. In the model without age, the parameter associated with N becomes not significantly different from zero and was removed. The equation with (dg, hdom and G) as regressors has a modelling efficiency value of 0.924 and a SSE value of 17844, values close enough to the previously found in more complex models. The final equation to be considered was:

(10) $$\mathrm{d_{med}} = \left(\mathrm{a1} + \mathrm{a2}\ \frac{\mathrm{G}}{1000}\right) \mathrm{dg}^{\mathrm{a3}}\ \mathrm{hdom}^{\mathrm{a4}}$$

III.4 Maximum diameter

The model for the maximum diameter has a structure similar to the one of the median diameter, a multiple allometric equation. The plots of the first stage (Figure 4) show that there is an evident relationship between d_{max} and dg, hdom, G and t, which in turn translates into being these variables that are present in the models with the lower values of SSE (Table A.3. in the appendices). There is an actual benefit in using dg and hdom rather than using just dg which translates into better values of modelling efficiency and lower values of SSE. The model with the best results is the one that uses (N, t and G) and that has an SSE value of 79045. The equation without age has a SSE value of 70243 and the same value of modelling efficiency than the previous one. The plots to analyse the regression assumptions and the plot of the actual versus predicted values were also very similar, showing that the presence of t does not have a great impact in the model. The final equation was then:

(11) $$d_{max} = \left(a1 + a2\ \frac{N}{1000} + a3\ \frac{G}{100}\right)\ dg^{a4}\ hdom^{a5}$$

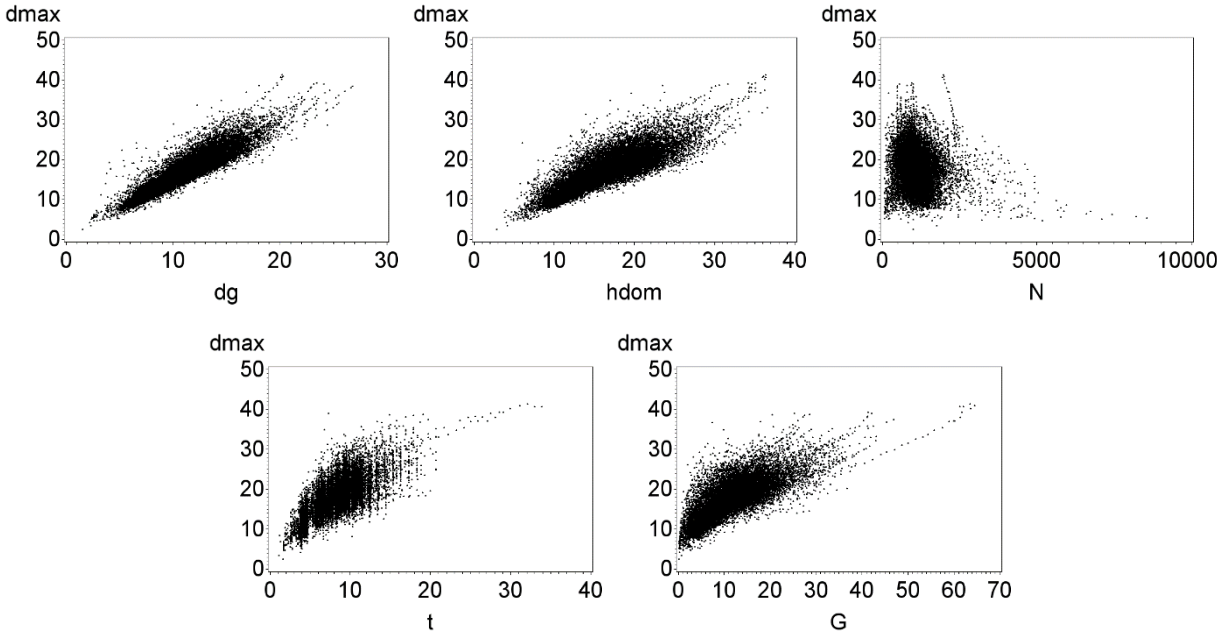


Figure 4. Plots of maximum diameter versus the different variables: quadratic mean diameter (cm), stand density (ha^{-1}), dominant height (m), age (years) and basal area ($m^2\ ha^{-1}$).

The selected functions all presented some heteroscedasticity, so the residuals were weighted, in an iterative process, with the inverse of the standard error of the estimate. The evidences of non-normal distribution of the model errors were corrected with robust estimation with the Huber´s function (Myers 1986).

The analysis of the autocorrelation plots for $\bar{d}$ indicated that it could be a first-order autoregressive model - AR (1), but when the model was adjusted, the AR(1) parameter turned out as not significatling differing from zero. The ARMA(1,1) was tested, since both AR and MA processes constitute good models for temporal series that can be observed in real life, but again the AR(1) parameter turned out as not significantly differing from zero, so the MA(1) structure of model errors was the one selected. For the remaining models, the analysis of the autocorrelation plots showed the typical patterns for a first-order autoregressive model - AR (1): the autocorrelation function (acf) rapidly decreasing to zero, the partial autocorrelation function (pacf) abruptly cuts after lag 1 and the value close to zero of the plotted pacf at lag 1 that means that the AR(1) model removes all the autocorrelation.

After the simultaneous adjustment, the hdom parameter in the d_{med} equation did not significantly differ from zero and was removed. The statistics used to characterize the prediction ability of the different equations fitted are in Table 3.

Table 3 – Summary of the statistics used to characterize the predictive capacity of individual models, after the simultaneous adjustment

	R^2_{adj}	SSE	MSE	MPRESS	MAPRESS
Minimum diameter	0.3736	26357.5	1.6157	-0.058727	0.90552
Average diameter	0.9935	958.6	0.0588	-0.13898	1.03042
Medium diameter	0.9558	7597.1	0.4657	0.044062	0.34280
Maximum diameter	0.8979	40522.4	2.4842	0.016882	0.16043

The variables $\bar{d}$ and d_{med} have very high values of modelling efficiency, while the value for d_{min} is much smaller, but similar results were found in the previous work (Mateus and Tomé, 2011). The values of average PRESS show that the models have small values of bias, and the precision, assessed by the average value of the absolute value of the PRESS residuals, is also very small if compared with the measurement errors that are expected for tree diameters. The box-plot graphics show also no relevant tendencies in the bias/precision of the models.

The QQ plots (not shown here) show that the final models have a close to normal distribution of the model errors, only d_{med} presents a small deviation more evident in the tails. No apparent evidence of error heteroscedasticity was shown when analysing the studentized residuals of the final models over the respective predicted values. The plots of the observed versus predicted values show a close to linear relationship with values close to the 1:1 reference line.

The graphic for d_{min} shows a considerable dispersion of the values, justifiable taking into account the smaller values of the other statistics.

The resulting equations that can be seen in Table 4 are different, in what concerns the combination of stand variables used as regressor, from the ones developed earlier by Mateus and Tomé (2011). The main structure of the equations is the same, but the regressor variables are different. The models were developed by those authors considering not only presumed relationships between the stand variables, but also in a way that observed and well-known biological processes where respected, so there was no reason to change the methodology that they designed. In what concerns the minimum diameter (d_{min}), this variable was modelled as a fraction of $\bar{d}$ instead of the minimum between $\bar{d}$ and d_{med}, $\bar{d}$ was removed from the parameter included in the exponential function and only stand level variables were selected.

Table 4 – Final equations

Model	Analytical expression
Minimum diameter (d_{min}, cm)	(12) $dmin = \bar{d} - \left(\bar{d} \ \frac{1}{1+\exp\left(a1+ a2\frac{N}{10000}+ \ a3\frac{coppice}{10}+ a4\frac{G}{10}\right)} \right)$ a1= - 0.45053 ; a2= - 0.68633 ; a3=2.92343 ; a4= - 0.30417
Average diameter ($\bar{d}$, cm)	(13) $\bar{d} = dg - \left(dg \ \frac{1}{1+\exp\left(a1+ a2\frac{N}{1000}+ a3\frac{G}{100}+ \ a4\frac{coppice}{10}\right)} \right)$ a1=3.059017 ; a2= - 0.19613 ; a3= - 0.37064 a4= 1.831762
Medium diameter (d_{med}, cm)	(14) $dmedian \ = \left(a1 + a2 \ \frac{G}{1000}\right) \ dg^{a3}$ a1=0.902281; a2= - 0.77368 ; a3= 1.038363
Maximum diameter (d_{max}, cm)	(15) $dmax \ = \left(a1 + a2 \ \frac{N}{1000} + a3 \ \frac{G}{100}\right) \ dg^{a4} \ hdom^{\,a5}$ a1=1.525803 ; a2=0.109678; a3=0.324231 ; a4=0.906173 ; a5=0.04426

The value of modelling efficiency for the d_{min} model was small when compared with the values obtained for the other models but it was similar to the one obtained by Mateus and Tomé (2011). These authors achieved a modelling of 0.44 and in our research the value was 0.37, but the amount of data used was much higher in this work and it covered many different situations, so a small decrease in the fit statistics is justifiable. Also, d_{min} is a variable that

depends from several factors, namely by tree mortality that depends on many factors. In fact, trees may be dominated during several years and just die in a particular year when the climate has some adverse events. $\bar{d}$ was modeled using the quadratic mean diameter - dg - as a basis, to ensure that the value is always lower than dg and the fraction to be subtracted from dg is a function of stand variables. The variables d_{med} and d_{max} were modelled both as multiple allometric relationships between stand characteristics.

Conclusion

The system of equations provided in this research includes regression models for minimum, average, median and maximum diameter distribution built in a way to obtain estimates that observe logical relationships according to biological principles. The system of equations can be used to simulate the diameter distribution when only stand variables are available.

The diameter variables predicted using the system of equations developed in this work, in connection with the variables basal area, dominant height, stand density and quadratic mean diameter that are usually available in the output of whole stand models, can be used in the Visual Basic program developed by Marto *et al.* (2009) that is implemented in the sIMFLOR interface (Faias *et al.*, 2012) in which the 3-PG is also implemented. This program can be used to estimate the parameters of a Johnson's SB distribution using a methodology similar to the one proposed by Parresol (2003). Using this functionality, the 3-PG and GLOBULUS models can add the diameter distribution to the set of variables that they provide as output.

Acknowledgments

This paper is a part of the PhD project of the first author, which is funded by a scholarship (SFRH/ BD/ 46594/ 2008) granted by Fundação para a Ciência e Tecnologia (Portugal).

References

Burkhart, H.E., Tomé, M. 2012. Modelling forest trees and stands. Springer, Berlin. pp. 261-298.

Faias, S.P., Palma, J.H.N., Barreiro, S.M., Paulo, J.A., Tomé, M. 2012. Resource communication. sIMfLOR – Platform for the Portuguese forest simulators. Forest Systems 21(3): 543-548.

Fonseca, T.F., Marques, C.P., Parresol, B.R. 2009. Describing Maritime Pine Diameter Distributions with Johnson's SB Distribution Using a New All-Parameter Recovery Approach. For Sci 55(4), 367-373

Furtado, A.M.X., 1997. Modelação da distribuição de diâmetros em povoamentos de eucalipto. Master Thesis. Universidade Técnica de Lisboa, Instituto Superior Técnico, Lisbon.

Hafley, W. L., Schreuder, H.T.. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research 7:481-487.

Hyink, D.M. 1980. Diameter distribution approaches to growth and yield modelling. In Forecasting stand dynamics (K.M. Brown and F.D. Clarke, eds.), p. 138-163. Sch For, Lakehead Univ, Thunderbay, Ontario.

Hyink, D.M., Moser, J.W.. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science 29:85-95.

Landsberg, J.J., Mäkelä, A., Sievänen, R., Kukkola, M., 2005. Analysis of biomass accumulation and stem size distributions over long period in managed stands of *Pinus sylvestris* in Finland using the 3-PG model. Tree Physiol. 25(7), 781-792.

Marto, M., Palma, J., Mateus, A., Tomé, M., 2009. Computer program for estimation of Johnson´s S_B parameters using a parameter recovery method. Publicações Científicas Forchange PC-1/2009. Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa.

Mateus, A., Tomé, M., 2011. Modelling the diameter distribution of eucalyptus plantations with Johnson's probability density function: parameters recovery from a compatible system of equations to predict stand variables. Ann. For. Sci. 68(2), 325-335.

Myers, R. H., 1986. Classical and Modern Regression with Applications. Second edition. Duxbury Classic Series.

Palahí, M., Pukkala, T., Trasobares, A., 2007. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). European Journal of Forest Research 126 (4), 563-571.

Parresol, B., 2003. Recovering parameters of Johnson's SB distribution. Research paper. USDA For. Ser., Southern Research Station, Asheville, NC.

SAS Institute Inc. 2009a SAS/ETS ® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 2876 p.

SAS Institute Inc. 2009b SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc. 7886 p.

SAS Institute Inc. 2009c SAS ® 9.2 Output Delivery System: User's Guide. Cary, NC: SAS Institute Inc. 984 p.

Schabenberger, O. and Pierce , F.J., 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press LLC 2000. Boca Raton, FL. pp. 211-213 and pp. 539-545.

Strub, M.R. and H.E. Burkhart. 1975. A class-interval-free method for obtaining expected yields from diameter distributions. Forest Science 21:67-69.

Appendices

Table A.1. – Residual sum of squares (SSE) for the models tested for d_{min} and $\bar{d}$ using alternative combinations of regressors.

	dmin				$\bar{d}$	
	$\bar{d}$		dmedian			
R2 base	0.115		0.018		0.983	
R2 full	0.358		0.278		0.986	
base	37258.8		41333.6		3462.3	
Full model	27041.6	- , - , +, - , -	30376.9	- , - , +, - , -	2986.1	- , - , + , - , +
N	35393	-			3320.9	-
hdom	32921.3	-			3452.4	+
coppice	33329	+			3420.5	+
t	33249.4	-			3413.4	-
G	30172	-			3460.4	+
N, hdom	31844	-, -			3305.6	- , +
N, coppice	30660	- , +			3273.3	- , +
N, t	31472.7	-, -			3348.9	- , +
N, G	30123.7	- , -			3276.8	- , -
t, G	29955.56	- , -			3275.3	- , +
hdom, coppice	29573.2	-, +			3406.2	+ , +
hdom, t	32305.9	-,-			3291.1	+ , -
hdom, G	30107.9	+ , -			3447.7	+ , -
G, coppice	27480	- , +			3415.5	+ , +
t, coppice	30829.4	-, +			3387.6	- , +
N, hdom, G	30092.2	- , + , -			3265.1	- , - , +
N, coppice, G	27148	- , + , -			3206.8	- , + , +
N, t, G	29777.9	- , + , -			3019.3	- , - , +
hdom, coppice, G	27457.2	+ , + , -			3404.4	+ , + , -
hdom, t, G	29724.9	+ , + , -			3291.2	+ , - , +
coppice, t, G	27466.9	+ , + , -			3326.4	+ , - , +
N, hdom, coppice	27858.3	-, - +			3252.3	- , + , +
N, t, coppice	28453.6	-, - +			3244.9	- , - , +
N, hdom, t	30962.6	-, - , -			3142	- , + , -
hdom, coppice, t	29470	- , + , +			3275.7	+ , + , -
N, hdom, coppice, t	27656.9	- , - , + , -			3121.4	- , + , + , -
N, hdom, coppice, G	27135.2	- , - , + , -			3190.2	- , - , + , +
N, hdom, t, G	29644	- , + , - , -			3019.1	- , + , - , +
N, coppice, t, G	27041.6	- , + , - , -			2986.2	- , + , - , +

Notes:

- the columns with signs indicate the signs of the variables included in the expansion of the *a* parameter
- the full model used N, hdom, coppice, G and t as parameters
- the shaded values of SSE indicate that one of the variables is not significant

Table A.2. – Residual sum of squares (SSE) for the models tested for d_{med} using alternative combinations of regressors.

dmedian	dg		hdom		dg, hdom	
R2 base	0.924		0.567		0.924	
R2 full	0.926		0.769		0.926	
base	17964		102198		17847	
Full model	17555	+ , - , - , -	54646	+, - , - , -	17496.8	+ , - , - , -
N	17963.3	-			17799	+
coppice	17962.9	+			17844.4	+
t	17641.9	-			17620.5	-
G	17900.7	-			17843.7	-
N, G	17775.3	+ , -			17668.9	+ , -
coppice, G	17900	+ , -			17841.5	+ , -
t, G	17641.6	- , -			17615.6	- , +
N, coppice	17692	- , +			17793.9	+ , +
N, t	17620.8	+ , -			17546.4	+ , -
t, coppice	17638.7	- , -			17618.9	- , -
N, coppice, t	17617.6	+, - , -			17545.9	+, - , -
N, coppice, G	17774.4	+, + , -			17665.1	+ , + , -
N, t, G	17557	+, - , -			17497	+ , - , -

Notes:

- the columns with signs indicate the signs of the variables included in the expansion of the *a* parameter
- the full model used N, coppice, t and G as parameters
- the shaded values of SSE indicate that one of the variables is not significant

Table A.3. – Residual sum of squares (SSE) for the different models for dmax using alternative regressors.

dmax	dg		hdom		dg, hdom	
R2 base	0.811		0.662		0.835	
R2 full	0.848		0.793		0.851	
base	100103		179102		87558.5	
Full model	80707.1	+ , + , + , +	109803	+ , + , + , +	79015.5	+ , - , +, +
N	89913.4	+			85779.4	+
coppice	99770.8	-			87009.5	-
t	85498.8	+			80248.3	+
G	88377.8	+			84807.7	+
N, G	88227.5	+ , +			80807.3	+ , +
coppice, G	88152.6	- , +			84413.9	- , +
t, G	80898.5	+ , +			79097.3	+ , +
N,coppice	89735.6	+ , -			85304.5	+ , -
N, t	81144.7	+ , +			79243	+ , +
t,coppice	85495.6	+ , +			80197.2	+ , -
N, coppice, t	81143	+ , + , +			79218.7	+ , - , +
N, coppice, G	88019.9	+ , - , +			84413.5	- , - , +
N, t, G	80707.2	+ , + , +			79045	+ , + , +

Notes:

- the columns with signs indicate the signs of the variables included in the expansion of the *a* parameter
- the full model used N, coppice, t and G as parameters
- the shaded values of SSE indicate that one of the variables is not significant

Chapter 4 - Modelling the 3-PG fertility ratio as a function of available soil water and other soil parameters

Executive summary

Local models to predict FR from site characteristics in eucalyptus stands in Portugal were developed. A general model was also developed using two different approaches: partial least squares and simple linear regression.

Abstract

- Context: Process based models are useful scientific tools, but have not been used very often as tools in forest management due to the requirement of detailed input data and the low detail and diversification of the outputs they usually provide. The 3-PG (Physiological Principles in Predicting Growth) is a physiologically based model developed with the intention of bridging the gap between the simpler empirical models and the more complex physiological based ones. It uses readily available data as input and provides stand level information with interest for the forest management. Soil fertility is taken into account through a fertility index, the so-called FR (fertility ratio), a value that varies between 0 and 1, empirically assigned according to the soil site characteristics. It is a parameter with a large impact in the model predictions, because changes in the FR affect the biomass production and the carbohydrates allocation for the roots by allocating a larger proportion of carbohydrates to the roots in poor sites. Although there is a limited knowledge of the relationship that exists between FR and soil properties some efforts have been made to develop empirical local functions relating FR and soil characteristics. Stape *et al.* (2004) developed a soil fertility response index using a linear combination of the amount of potassium, phosphorus and cation exchange capacity. Almeida *et al.* (2010) presented FR as function of soil fertility and available soil water, considering also other factors such as slope and management effects. Vega-Nieva *et al.* (2013) developed a model to predict the value of FR considering the available soil water and other nutrients with significant effect on fertility such as phosphorus, potassium, calcium and magnesium.

- Aims: The present study aimed at developing a general model to predict FR from soil characteristics. An additional objective was to find which nutrients are the most limiting for eucalyptus growth in Portuguese stands and test the possibility of making the 3-PG model sensitive to fertilizations.

- Methods: Data came from 5 fertilization trials located along the most important areas of eucalyptus in Portugal. For each one of the sites, a local model was developed to predict FR from soil characteristics using 2 different methodologies: partial least squares and simple linear regression. After that, a global model was also developed.

- Results: The models developed for each one of the 5 sites showed a good performance (with R^2 between 0.24 and 0.97, small bias and high precision) but the generalized model did not achieve such a high predictive ability (R^2=0.34).
- Conclusion: The results show that fertilization must be locally adapted and indicate the need to broaden this analysis to more sites.

Key-words: 3-PG, fertility ratio, soil characteristics, nutrients, fertilization.

Introduction

Prediction of future forest development under alternative management is essential for forest planning. This has traditionally been achieved using the so-called empirical growth and yield models. Empirical models describe forest development with growth functions fitted with large data sets through statistical regression techniques. They are commonly used for forest management and planning, because they supply correct and reliable forest productivity information (Levy *et al.*, 2004), although its use is restricted to the site conditions and forest management practices for which there were growth data available in the fitting data set. Most empirical models make use of the concept of site index (S), an indirect measure of the potential growth at a site assessed by the expected dominant height at a reference age. Implicit in this concept are the assumptions that the dominant height growth of a stand is independent of changes in the environment and is not influenced by stand stocking. When growth conditions are stable, one can use this concept without problems, but when the interest falls on new locations, changing soil or climatic conditions, the growth response is highly non-linear and the relationships can be very difficult to predict. The 3-PG (Physiological Principles in Predicting Growth) model (Landsberg and Waring, 1997; Sands and Landsberg, 2002) is a physiologically based model developed by Landsberg and Waring (1997) with the intention of bridging the gap between the simpler empirical models and the more complex physiological based ones. It simplifies some of the processes and was designed to require just readily available data as input. Other than just the biomass production it also provides as output stand level information with interest for the forest management (volume under bark and the respective annual increment, stand basal area and quadratic mean diameter). It has the ability to reflect the effect of changes in the environment, is well documented and the code is available. The research group in which the present research was conducted has been working on this model and it has already been calibrated for Eucalyptus plantations in Portugal (Fontes *et al.*, 2006).

It is well known that the plant nutritional status has a significant effect on productivity and on allocation to roots. Landsberg and Waring (1997) emphasized that although there is evidence to suggest that nutrition affects the amount of carbon allocated to roots, the quantification of this relationship is difficult. Furthermore, intensive forestry has shown that increases in productivity can be achieved improving soil nutrition, because an increase in fertility increases the leaf area index (LAI). The 3-PG model uses a simplified method to express soil fertility, through a fertility index, the so-called FR (fertility ratio), a value that varies between 0 and 1 empirically assigned according to the soil site characteristics. The value 0 corresponds to very poor sites and the value 1 for sites at which nutrients are not limiting growth. Landsberg *et al.* (2003) suggested the use of soil studies to estimate FR, but this method is not practical and

has many limitations. In the same study, the suggestion was made to introduce a dynamic model of soil organic matter and this suggestion has been followed in several subsequent studies with the integration of 3-PG with models of soil organic matter decomposition (Peng *et al.*, 2002; Hirsch *et al.*, 2004; Paul and Polglase, 2004), although none has achieved an integration that allows the FR to be estimated from representations of soil processes. Although there is a rather limited knowledge of the relationship between the FR and the soil properties (Landsberg *et al.* 2003), efforts have been made in linking nutrient cycles with 3-PG, as well as advances in the development of local empirical functions between FR and soil characteristics. Stape *et al.* (2004) made a critical analysis using paired plots in eucalyptus plantations across 40 sites that captured the regional soil and productivity differences. The paired plots were initially similar (the same age and similar starting values of woody biomass) and one of them was fertilized to eliminate any limitation on growth. There were two growing seasons considered in the study, one with a normal climate and the other one was very wet. A fertilizer response (FER, t ha-1) that measures the response of a site to fertilization was determined from the observed growth data for the wet years. 3-PG was also applied to the paired sites and the value of FR adjusted so the predicted wood increment matched the observed increment in the wet years. When these values where compared with the observed FER, it was found that FR≈1 for sites with no fertilizer response (FER≈0) and that FR declined with increasing FER. The relationship between FER and soil parameters was also observed and a soil fertilizer response index (SFRI) was built, considering a linear combination of extractable potassium, phosphorus and cation exchange capacity in the upper soil layer that explained 56% of the observed variation in FER, but it is not known if the this SFRI relationship is general. The study demonstrated that FR and the manner in which site fertility effects are included in 3-PG, allow the prediction of observed responses to site fertility. Almeida *et al.* (2010) presented a model of FR as a function of soil fertility and available soil water, also considering other factors such as slope and management effects. In the work of Vega-Nieva *et al.* (2013), several plots were selected on which the soil characteristics were studied, including a full description of the soil profile, texture, pH, organic matter and fertility analysis. For each plot, the value of the available soil water was calculated and the value of FR was determined by minimizing the differences between the observed and estimated biomass values. A model for predicting the value of FR from the amount of available soil water and other nutrients with significant effects on fertility, in this case phosphorus, potassium, calcium and magnesium, was developed.

Plants require a number of mineral nutrients for their growth and inadequate supplies will affect growth. Nitrogen is ultimately derived from atmospheric nitrogen, but all the other nutrients have their origin in the soil material. The macro-nutrients: nitrogen (N), phosphorus (P), sulphur (S), calcium (Ca), magnesium (Mg), and potassium (K) are required in relatively large

quantities. The micro-nutrients or trace elements are boron (B), chlorine (Cl), cobalt (Co), copper (Cu), iron (Fe), manganese (Mn), molybdenum (Mo), nickel (Ni), and zinc (Zn). The current ability to describe soil nutrient status in terms that can be used in management-oriented models is very limited and even 3-PG cannot be used to give guidance on fertilization needs of eucalyptus stands (Landsberg *et al.*, 2003; Landsberg and Sands, 2010; Sands, 2003). The improvement of the 3-PG with a model that predicts FR from soil characteristics is important not only because the empirical way in which FR has to be estimated is recurrently appointed as a drawback of the model, but also because it will make 3-PG sensitive to fertilizations and to find what are the most limiting nutrients in eucalyptus production in Portugal.

These were then the main objective of this study: to analyse and develop a generalized model to predict FR from soil characteristics. Additionally the study must also contribute to find which nutrients are the most limiting for eucalyptus growth in Portuguese stands and to check to which nutrients the growth shows better responses. The achievement of these objectives will make 3-PG able to simulate the effect of fertilization on eucalyptus productivity, a really important result with practical implications.

Data

The work was conducted in 5 different properties belonging to the pulp company ALTRI in which the company had established as was monitoring fertilization trials. There was a concern in trying to consider different soil types and climatic areas. The properties are scattered along the Eucalyptus distribution area in Portugal (Figure 1).

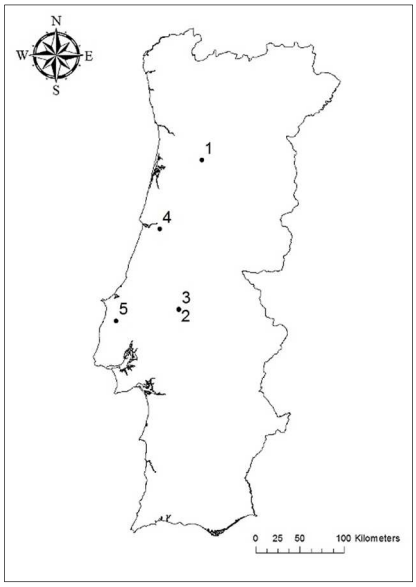


Figure 1. Trial distribution.

All plots are of first rotation stands of *Eucalyptus globulus*, non-clonal, planted at a 4 m x1.8 m spacing and considering the application of different fertilizers. Plots where measured from the age of 2 to 8.8 years, having each one at least 4 measurements. Each plot had an area of 518.4 m^2 with a double edge and a number of 12 or 16 central plants for height and diameter measurements. All trees were measured for diameter and height and the aboveground biomass and biomass per tree component (stem wood, stem bark, branches and leaves) were estimated using the equations developed by António *et al.* (2007). Crown length, used as regressor in the leaf and branches biomass equations, was estimated with the crown ration and crown length system of equations developed by Oliveira and Tomé (submitted).

On each site, different types of fertilization were tested alongside the traditional fertilization, Foskamónio 7-21-21 (7% nitrogen (N), 21% phosphorus (P2O5) and 21% potassium (K2O)) (Table 1). Each treatment was repeated several times and every treatment was repeated 6 or 4 times, so every site had a variable number of plots as can be seen in Table 1.

Table 1 – Fertilizers used in the different treatments of the studied sites

Site	Replicates	Treatment / Fertilization	
1	6	1	Foskamónio 7-21-21 (traditional fertilization)
		2	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4)
		3	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4) + Cu (SO4) + borax (7,2g)
		4	$CaCO_3$ + Osmocote
2	6	1	Foskamónio 7-21-21 (traditional fertilization)
		2	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4)
		3	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4) + Cu (SO4) + borax (7,2g)
3	6	1	Foskamónio 7-21-21 (traditional fertilization)
		2	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4)
		3	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4) + Cu (SO4)
		4	11-22-9+6 MgO (11% N + 21% P_2O_5 + 9% K_2O + 6% MgO)
4	6	1	Foskamonio7:21:21
		2	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl
		3	$CaCO_3$ + P_2O_5 18% + NO_3 + KCl + Zn (SO4) + Cu (SO4) + borax (7,2g)
		4	11-22-9+6 MgO (11% N + 21% P_2O_5 + 9% K_2O + 6% MgO)
5	4	1	Foskamónio 7-21-21 (traditional fertilization)
		2	Zinc sulfate + 280 Kg/ha Single superphosphate 18% (18% P2O5 + 16% Ca + 8% S)
		3	Zinc sulfate + 560 Kg/ha Single superphosphate 18%
		4	Zinc sulfate + Triple super phosphate 43% (43% P2O5)
		5	Zinc sulfate + 1120 Kg/ha Single superphosphate 18%
		6	Zinc sulfate
		7	Osmocote

The climatological data used, covering the growth period of the stands until the last measurement available, was in part collected in the SNIRH (Sistema Nacional de Informação de Recursos Hídrico) webpage (http://snirh.pt/) and part from the Instituto Português do Mar e da Atmosfera (IPMA). Data until 2009 was collected from SNIRH, all the data from 2009 and after came from IPMA. The network of climatological stations was analysed and the ones closer to the sites and with similar altitudes were selected. No altitude correction was made to the data due to the small differences between the sites and the stations used.
Table 2 displays the summary of the climatic conditions for the different sites under the time period that was considered in the analysis and stand age covered by that same period.

Table 2 – Average climatic conditions for the study sites during the period of analysis (2005-2012)

Site	Age	Tmin	Tmax	Rain	Rain days	Solar rad	Frost days
1	2-8.8	-2.9	32.1	552.6-1371.6	121-257	3842.3-6069.3	7-78
2	2-7.3	3.2	33.9	380.5-943.5	82-205	5231.9-5636.01	0-7
3	2- 7.3	3.2	33.9	380.5-943.5	82-205	5231.9-5636.01	0-7
4	2-8.6	2.2	31.2	608.8-936.3	125-241	5296.2-6046.15	0-14
5	2-6.1	6	30.4	412.8-884.2	125-208	5296.2-6046.15	0-2

Note: Age-period of stand growth under analysis (years); Tmin-minimum of monthly mean daily minimum temperature (ºC); Tmax-maximum of monthly mean daily maximum temperature(ºC); Rain- Total annual precipitation (mm); Rain days- Total annual days of precipitation (days); Solar rad- Total annual solar radiation (MJ/m^2/year); Frost days-Total annual frost days

All the plots considered have at least 4 inventory measurements along the growth period, which ranged from 6 to 9 years old and cover a wide range of stand characteristics, as can be seen in Table 3.

Table 3 – Summary of the stand variables of the selected study sites

Site	plots	n	Age (years)	hdom (m)	G (m^2/ha)	N (ha^{-1})	S (m)
1	24	6	2.2-8.8	6.6-22	1.63-24.85	1389	18.45-25.49
2	18	4	2.1-7.3	4.3-20	0.49-15.47	434-1389	17.82-23.69
3	24	4	2.1- 7.3	3.5-17.4	0.35-13.27	608-1389	16.84-23.29
4	24	4	2-8.6	1.9-23	0.49-23.38	694-1389	13.68-25.12
5	28	4	2.8-6.1	5.4-19.4	2.11-21.51	1042-1389	17.05-28.35

Note: plots – number of plots considering the different treatments and repeats; n – number of measurements made on each plot; Age-period of stand growth under analysis (years); hdom-dominant height (m); G-basal area (m^2 ha^{-1}); N- stand density (ha^{-1}); S- Site Index (m)

Fertilization was applied at planting and soil analysis was made in each plot before the planting procedure and around the second year after the planting. In each plot, 6 simple soil samples were collected on the depths of 0-15cm and 15-30cm. The simple samples were mixed and one general sample was formed containing 2 parts of the 0-15cm sample and 1 part of the 15-30cm sample. Physical parameters (percentages of sand, silt and clay) were determined in each sample as well as the following chemical parameters: pH, organic matter percentage, Egner-Riehm P, K, Ca, Mg, Cu, Fe, Mn, Zn and B.

On each of the properties, some plots were selected as representative of the soil types present and used to conduct a complete soil profile. On the representative plots selected, a soil pit was dug and the profile completely described by horizons, including texture, bulk density and water retention capacity. For every horizon a soil sample was collected for further laboratory analysis. First the samples where dried at 105ºC and then sieved for coarse fragments (> 2 mm) and fine earth (< 2 mm). The fine earth fraction was then analysed for sand, silt and clay percentage (Bouyoucos densimiter/hydrometer method) and organic C (g/Kg) that was used to calculate the organic matter content. Each plot in the property was assigned to one of the soil pits so that the soil characteristics of each plot needed to run the 3-PG model were available.

The maximum available soil water was estimated for the whole profile in the field (ASW_{field}) and also calculated using the Domingo Santos *et al.* (2006) expression that considers soil texture, organic matter, fine earth percentage and slope (ASW_{calc}):

$$ASW_{calc} = \left[12{,}5\ h_e + 12{,}5\ (50 - h_e)\ \frac{k}{2}\right]\ c\ \frac{TF}{100}$$

with h_e – moisture equivalent of the horizon's fine earth; k – permeability dependent parameter; c – slope related variable; TF – fine earth percentage of the horizon. For more details on the parameters and calculation steps, please check Domingo Santos *et al.* (2006). The value is calculated for each of the horizons of the soil profile and then a global value is estimated for the soil considering the partial horizon values and respective horizon depth.

The calculated and field determined ASW and the soil properties for the selected sites are described in Table 4.

Table 4 – Soil properties and estimated and calculated available soil water for the sites

Site	Parent material	Soil texture	Depth (cm)	C (%)	CF (%)	ASW_{calc} (mm)	ASW_{field} (mm)
1	Schist	Sandy loam	40-310	0.6-4.7	11.2-49.9	20-206	21-190
2	Sandy materials	Sandy loam	95-200	3.7-17.2	17.4-30.4	141-147	75-86.2
3	Sandy materials	Sandy loam	150-200	5.8-21.7	18.2-29.6	300-412	61.8-99.8
4	Sandy materials	Sandy loam	170-210	2.4-8.6	1.6-18.5	230-264	225-328
5	Sandstone	Sandy loam	120-300	1.9-42	1.5-9.5	67-214	101-196

Note: C- carbon content; CF- coarse fragments; ASW_{calc}- available soil water calculated with the Domingo Santos at al. (2006) expression; ASW_{field}- available soil water estimated in the field

Methods

III.1 FR estimation

The 3-PG implemented in Excel (Visual basic for applications), version 2.7 was used (as available at http://booksite.elsevier.com/9780123744609/?ISBN=9780123744609) with the parameter values for eucalyptus stands in Portugal from Fontes *et al.* (2006).

After estimating the ASW values for each of the plots, the corresponding FR value was determined by comparison of observed and estimated biomass values. Within each plot, the FR was assumed not to vary with time. To find this value, the climatic data of the growth years was used, alongside the values of ASW_{field} and ASW_{calc} (one at a time). The FR value was allowed to change until the value that originated estimated biomasses closer to the observed ones was found. In order to do that, not only the plots of the observed and estimated values of biomass were observed, but also the mean squared errors (MSE), precision and bias were calculated.

III.2 FR modelling

The total soil content of each nutrient (in % or kg/ha) was calculated considering the depth and soil bulk density. Following the work of Vega-Nieva *et al.* (2013), the total contents of C, N, P, K, Na, Ca and Mg was divided by the asymptotic value of the nutrient for plant growth. These values were assumed as the contents above which tree growth limitations are not expected, based on the study provided by Pérez Cruzado *et al.* (2011).

Water and nutrient absorption are 2 independent processes, but they are connected. The plants need water not only to grow, but also to be able to absorb nutrients from the soil, so a series of variables obtained by multiplying the nutrient values by the water content was tested. These variables (nutrient_a) represent the water-nutrient interaction.
There were soil analysis available from 2 moments in time, prior to the stands establishment and after the fertilizations were made. Since the growth response of the stand is related to the amount of extra nutrients supplied by the fertilization, variables representing the difference of the amount of each of the nutrients in the soil were also calculated (nutrient_dif) and the results from both soil analysis where tested in the modelling process. The “nutrient” variable relates to the soil analysis made prior to the stand establishment; the “nutrient_2013” variable to the soil analysis made at the time of the soil pits opening and the “nutrient_dif” variable to the difference between the two periods.

The number of observations (118) is small when considering the number of explanatory variables, so the analysis was conducted in two stages. On a first stage, an exploratory analysis was made on each of the sites to see which variables related better to the FR. Two different approaches were used, simple linear regression and a partial least squares procedure (PLS). For that purpose, the PROC REG and PLS procedures of SAS (SAS Institute Inc., 2009) were used.

The REG procedure is one of the many regression procedures available in the SAS System. It is a general-purpose procedure for regression that allows not only to adjust simple linear regressions, but also to select models considering numerous candidate variables. First, an analysis was undertaken considering all the possible variables according to a RSQUARE selection method that finds the models with the highest values of R^2 in the range of all possible models with 1,2, 3,…, k regressors (k is the total number of variables). The models with the highest values of R^2 and appropriate number of parameters were then adjusted with the REG procedure to ensure that all the parameters were significantly different from zero and the best model for each site was selected.

Due to the high number of possible variables, the relative small number of observations and to allow the comparison with the work of Vega-Nieva *et al.* (2013), the PLS procedure was also used to fit the models. PLS analysis is recommended for data sets with large numbers of independent variables compared with the number of observations and for data sets where there is a possible self-correlation between the independent variables (Wold, 1995). While ordinary least squares regression has the single goal of minimizing sample response prediction error, seeking linear functions of the predictors that explain as much variation in each response as possible, the PLS procedure has the additional goal of accounting for variation in the predictors, under the assumption that directions in the predictor space that are well sampled

should provide better prediction for *new* observations when the predictors are highly correlated. All of the techniques implemented in the SAS PLS procedure work by extracting successive linear combinations of the predictors, called *factors* (also called *components*, *latent vectors*, or *latent variables*), which optimally address one or both of these two goals—explaining response variation and explaining predictor variation. In particular, the method of partial least squares balances the two objectives, seeking factors that explain both response and predictor variation.

The PLS results were interpreted using the VIP (variable importance in the projection) scores and considering as criteria for interpreting the significance of the variables a VIP > 0.8 (Wold, 1995). The models were also evaluated considering the value of the prediction R^2, the value of R^2 calculated with the sum of the PRESS residuals that resulted from the one-at-a-time cross validation method used:

$$R^2{}_{pred} = \frac{PRESS}{SST} \quad , \quad \text{with} \quad PRESS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_{i,-i}\right)^2$$

On the second stage, a global model for all the sites was adjusted using both PLS and PROC REG. The model evaluation was made considering the values of the R^2 and also some statistics based on the PRESS residuals (residuals computed by a jack-knife procedure that indicates the predictive ability of the equations by cross validation (Myers, 1986)):

- model fitting capacity, $R^2{}_{pred}$ as defined above;
- mean value of the PRESS residuals, MPRESS (model bias);
- mean absolute value of the PRESS residuals MAPRESS (model precision).

Results

IV.1 FR estimation

The difference between using the available soil water calculated with the Domingo Santos *et al.* (2006) expression or the available soil water estimated in the field with the analysis of the horizons in the soil pit and its characteristics is very small (Table 4). The differences show that it is possible to have very good results, compatible with the ones from the field, with a simple expression based on soil texture, organic matter, fine earth percentage and site slope.

The 3-PG was run with both ASW_{calc} and ASW_{field} and the difference in the 3-PG results from using one or the other was almost none. It was found that the model, in the studied conditions, has little sensitivity to small differences in the ASW value used, even when ASW values are small. On the other hand, the model revealed to be very sensitive to the initial values of biomass, especially the initial value of leaf biomass. Matala *et al.* (2003) had already found that PBMs, compared to statistical models, are much more sensitive to initial stand conditions

and silvicultural treatments. Small changes in the value of initial leaf biomass on the 3-PG model led to very different growth rates under the same conditions. To further test this influence and to compare it with the influence of available soil water in the growth, some model runs were made increasing the water supply with additional irrigation and by changing the initial values of biomass. Both changes have a positive effect in the growth speed of the stands, but the model is much more sensitive to changes in the initial biomass values.

The summary of the adjusted FR values with the respective values of MSE, bias, precision and ranges of the values of the soil characteristics in each of the sites can be seen in Table 5.

Table 5 – Ranges in soil texture percentages, nutrient concentrations and optimized FR values derived by the best fit of the plots analysed

Site	1	2	3	4	5
Sand (%)	75.9-88.2	77.3-78.2	63.3-73.3	64.7-79.2	81-99.2
Silt (%)	7.4-13.3	9.2-12.9	7.9-9.9	7.9-19	4.9-9.5
Clay (%)	3.9-10.2	9.8-12.6	16.8-28.8	9-16.3	4.8-24.7
N (%)	0.3-0.5	0.1-0.3	0.1-0.3	0.1-0.2	0.1-0.2
P_2O_5 (kg/ha)	0.7-3.9	0.3-3.8	0.7-3.8	0.3-2.1	0.9-4
K_2O (kg/ha)	8.9-25	19.5-52.8	25.6-54	11.5-21.4	12.6-51.3
K (kg/ha)	9.8-24.6	18.6-43.8	19.6-46.9	8.9-22.0	5.5-20.9
Ca (kg/ha)	0-97.4	62.4-222.9	49.1-226.8	21.1-93.6	8.1-158.2
Mg (kg/ha)	1-8.2	6.4-29.7	10-37.4	4.1-14.9	3.2-38.7
Na (kg/ha)	2.2-8.2	1.8-6	4.7-11.7	1.4-3.4	2.9-22.6
Zn (kg/ha)	0-0.2	0.1-0.3	0.1-0.4	0.1-0.3	0.9-2.4
Cu (kg/ha)	0-0.1	0.1-0.3	0.1-0.3	0.1-0.3	0.1-0.4
Mn (kg/ha)	0.2-0.7	0.5-1.8	0.8-2.7	0.5-3	0.7-3.1
Fe (kg/ha)	16-39.7	21-88.2	13.8-96.4	23.3-56.8	27.7-92.5
B (kg/ha)	0.1-0.3	0.1-0.4	0.1-0.4	0-0.2	0.2-0.7
Optimized FR	0.01-0.8	0.01-0.5	0.01-0.6	0.01-0.5	0.01-0.5
MSE	0.62-5.52	0.19-9.87	0.09-16.11	0.40-8.80	0.33-30.67
Precision	0.05-1.90	0.05-2.3	0.05-2.95	0.03-2.51	0.01-2.82
Bias	-1.41-1.90	-2.3-1.82	-2.95-1.14	-2.51-1.33	-2.82-1.69

Note: C- carbon content; CF- coarse fragments; ASW- available soil water calculated with the the Domingo Santos at al. (2006) expression; Water- available soil water estimated in the field

The soil presents differences among plots inside each of the sites, as expected due to the large spatial variation of the soil properties that is characteristic of most sites in Portugal. Different types of fertilizers containing different nutrients were used in the different plots originating differences in the soil content in nutrients (Table 1). The Foskamónio 7-21-21 is the

traditional fertilization and was used as the control. The values obtained for FR are very variable within each of the sites. Even though the plots were located next to each other, the growth variation of the trees was very high. Soil characteristics are very variable and even with the repeated treatments, it is very difficult to eliminate all the possible differences between plots and reach a medium representative value.

Nitrogen (N) is the only nutrient that has no common soil mineral source, so nitrogenous fertilizers are an important source of N. It is the nutrient needed in greatest amounts by plants and is important as a component of proteins, so N supply is frequently used as the reference element in considering plant nutrient uptake and dynamics. In the forms useful to plants, nitrogen is probably the nutrient most universally limiting plant growth.

The sites presented extremely low phosphorus (P) content, with a high potential response to P fertilization. Organic matter and potassium (K) contents were medium to low, with some potential response to N and P application, so it makes sense that the traditional fertilization, that has always shown some results, is a NPK fertilizer.

The different fertilizations tended to lead to higher values of FR (Figure 2). Treatment 1 is the traditional fertilization (Table 1) and the other treatments tend to have better results, with some exceptions. There are very low values of FR in some plots in all the sites, but after a careful analysis, a single explanation for the bad growth rate was not found. Some plots have higher values of some nutrients, giving some indication of toxicity, for example in site 1, treatments 2 and 4 (Figure 2).

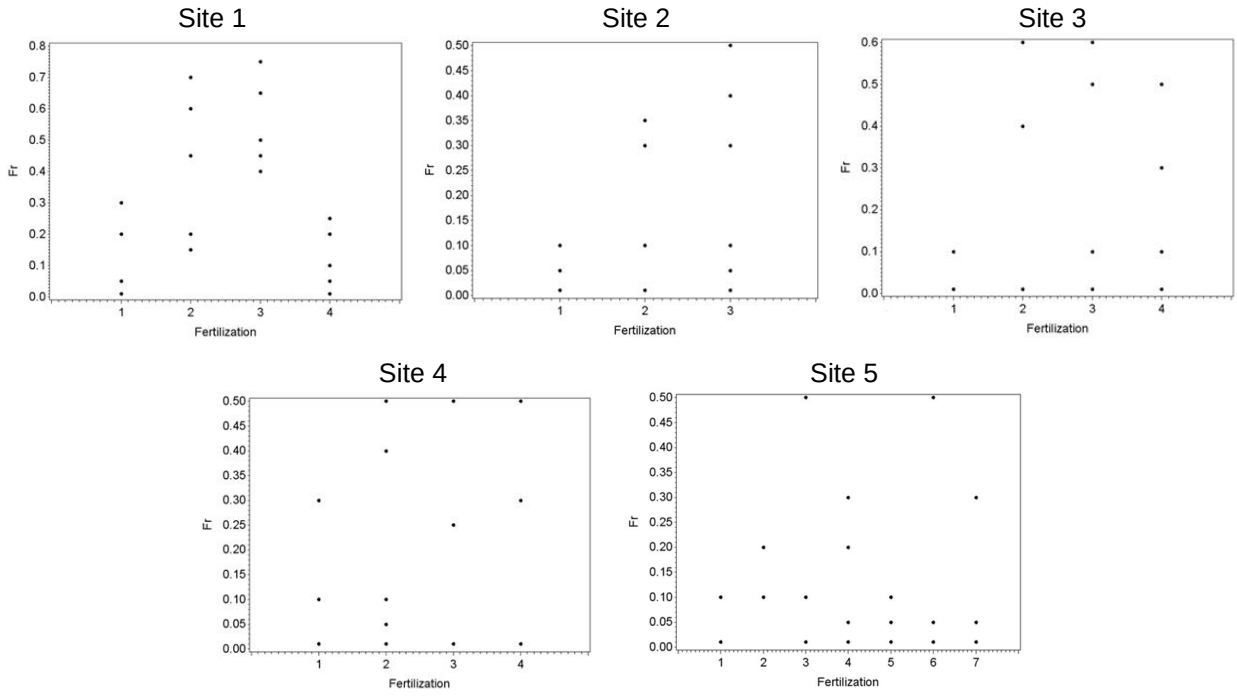


Figure 2. Plots of adjusted FR versus the treatment for each of the sites.

The 2 low values of FR in treatment 2 have higher values of P_2O_5 than all the other, while the 3 lower values in treatment 4 have higher values of Fe and Na. Other points are related to the lack of nutrients, like the low values of site 2 treatment 1 – low values of B – treatment 2 – low values of Na – and treatment 3 – low values of Mg – or the low values of site 3 treatment 2 – low values of Zn.On site 1, the difference between treatment 1 and 4, the application of $CaCO_3$, did not seem to improve fertility and the change of the traditional Foskamónio with a similar fertilizer with slower release seems to have a negative effect on the growth. The other treatments have a clear effect on stand growth, with the additional nutrients supplied increasing the growth rate of the stand. Site 2 shows similar results, the treatments are the same as the ones used in site 1. The additional Zn, Cu and B supplied in treatments 2 and 3 have great impact on the FR value, indicating the importance of these micronutrients in tree growth. Site 3 shows similar results, treatments 2 and 3 are similar to the other sites, but no B was added. Both fertilizations present good responses, better than the one resulting from treatment 4, which indicates that maybe the response is due to Zn and/or Ca, the elements that are not present in fertilization treatment 4.

The results in site 4 are not so clear, besides the fact that the other fertilizations are definitely better in most plots than the traditional one. Site 5 has highly variable values of FR on each treatment, making the conclusions also difficult. The traditional treatment is the one with the lowest values of FR, similar to the ones from treatment 5. All the other treatments seem to raise the fertility, but the values are too dispersed to say which one is better.

IV.2 FR modelling

IV.2.1 Site-specific FR models

Table 6 shows the results of the adjusted models for each site using both procedures, PLS and PROC REG. The results from both procedures are different, PLS tends to have more explanatory variables and lower values of R-squared and the variables differ between the models for the same site.

In the work of Vega-Nieva *et al.* (2013), the ratio of the nutrients with the optimum value showed better results as explanatory variables that the nutrient values, but that was not the case here. In both procedures, these values did not appear to perform better than the nutrients content values themselves, the same result obtained also with the variables related to the differences in nutrients content before and after the fertilization treatments. On the other hand, the variables that represent the water-nutrient interaction appear in almost all the models, even if the available soil water does not always appear as explanatory variable.

Table 6 – Model variables and values of R^2 pred (PLS) and R^2 (PROC REG) for each site using the PLS and PROC REG procedures

Site	1		2		3		4		5	
	PLS	REG	PLS	REG	PLS	REG	PLS	REG	PLS	REG
ASW		X				X				X
Water							X		X	
Om	X			X			X		X	
Sand (%)							X		X	
Silt (%)							X		X	X
Clay (%)										X
N (%)										
P_2O_5 (kg/ha)	X							X		
K_2O (kg/ha)		X								
K (kg/ha)	X			X	X				X	
Ca (kg/ha)		X	X		X					
Mg (kg/ha)	X					X				
Na (kg/ha)	X	X	X	X					X	
Zn (kg/ha)					X	X				X
Cu (kg/ha)	X		X		X	X				
Mn (kg/ha)								X		
Fe (kg/ha)	X	X			X	X	X	X	X	X
B (kg/ha)										
N_a									X	
P_2O_5_a						X		X	X	X
K_2O _a										
K _a										
Ca _a			X						X	X
Mg _a						X				
Na _a										
Zn _a		X		X	X	X				
Cu_a	X		X	X	X				X	
Mn _a								X	X	X
Fe_a				X	X				X	
B _a								X	X	
Bases		X	X	X	X					
R^2 pred / R^2	0.35	0.75	0.32	0.97	0.25	0.76	0.24	0.55	0.46	0.71

Water is a controlling factor in the growth of trees and any attempt to model or explain the growth of forests must include calculations of the soil water balance. Maximum available soil

water appears only in some of the adjusted models, but, on the other hand, some of the variables are accounted for considering their interaction with water. Asides from that, 3-PG already takes into account maximum available soil water in the soil water balance sub-model, so the fact that this variable does not appear in all the models, does not necessarily means that stand fertility is not related or is not affected by soil water.

Trying to explain or understand why the different nutrients appear or not as explanatory variables on each of the sites is not a simple task. First, the 2 approaches lead to different variables and there is no evident relationship between the nutrients supplied by the different fertilizers and the nutrients that explain the FR differences. For example, on site 4 there is the traditional NPK fertilizer; a treatment with Ca, P, N and K; a treatment similar to the previous added with Zn, Cu and B and a treatment with N, P, K and Mg. It would be expected that Ca, Zn, Cu, B and/or Mg would appear as explanatory variables in the FR model, but there is only Fe in the PLS model and P, Mn, Fe and B in the linear regression model.

Asides from that, the local models tend to have good predicted capacity, but when comparing the models obtained for each local it is difficult to observe any global trend.

IV.2.2 Global FR model

After the fitting of the models for each of the sites revealed no group of nutrients that could be used for a global model, the total group of variables was again used to adjust a global model. A second model was also adjusted using only the plots with the traditional fertilization to see if the resulting models, without the influence of the different fertilizations, even if that meant fewer observations, performed better. The results of those models can be seen in Table 7.

The PLS procedure resulted in a model with numerous parameters and an R^2pred value of only 0.19. Considering only the plots with the traditional fertilization, the model has only 3 variables, organic matter, Mn and Mn_a, but an R^2pred value of 0.13. The variance importance plots (Figure 3) of the parameters show that all the variables are significant according to Wold´s criterion and the fit diagnostic plots (Figures 4 and 5) show what it can be seen in the numeric values, the models are not excellent, but have some predictive capacity.

The PROC REG procedure originated models with less variables and higher values of R^2 (Table 7). Some of the nutrients like Zn, B or even Ca were expected in the model. Zn and B are nutrients that appear in only some of the treatments and that might explain the differences in the FR values. N is a fundamental nutrient in tree nutrition and growth and the soils had low values of N, so its presence makes sense, even if it just appears related to water in the N_a form. Ca and Mg also appeared in the model developed in the work of Vega-Nieva *et al.* (2013) and Ca can play an important role in the assimilation of other nutrients.

Table 7 – Parameter estimates and statistics values using the PLS and PROC REG procedures for the global models with all the plots (left) and only with the traditional (treatment 1) treatment plots (right)

	PLS		PROC REG	
Intercept	-15.74630715	0.085753	-17.12622	-0.53807
Water	0.00315612			0.00161
Organic matter	0.10611682	0.005686	0.11326	0.11304
P_2O_5 (kg/ha)	0.07094055			
K (kg/ha)	0.00141749			
Mg (kg/ha)	-0.00764230			
Zn (kg/ha)	-0.45843812		0.13395	-0.43567
Cu (kg/ha)	-0.56428650			
Mn (kg/ha)	-0.01085313	-0.02552		
B (kg/ha)	1.23238258		0.33134	0.73297
Sand (%)	-0.00298714			0.00989
Silt (%)	-0.00321782			
Clay (%)	-0.00158110			-0.01297
N_a	-0.00825786		-0.00889	-0.01066
P_2O_5_a	-0.00014543			
Ca _a	0.00000358			
Zn _a	0.00289296		0.00341	
Cu_a	0.00273278			
Mn _a		-0.00005615		
Fe_a	-0.00000424			
B _a	-0.00189529			
P_a	-0.00032463			-0.00018497
K_a				0.00002193
Bases			0.00305	
Bases min			-14.54870	
Days of frost	-0.02955115		-0.02100	-0.03632
Solar radiation	0.10855991		0.10813	
R^2 pred / R^2	0.19626	0.13047	0.20415 / 0.3407	0.35054 / 0.5850
MPRESS	0.00109	0.00017	.00020677	0.0056878
MAPRESS	0.14121	0.054865	0.14002	0.067657

Note: Bases=Ca+K+Na+Mg; Basesmin – the minimum value of (Camax, Kmax, Namax, Mgmax);

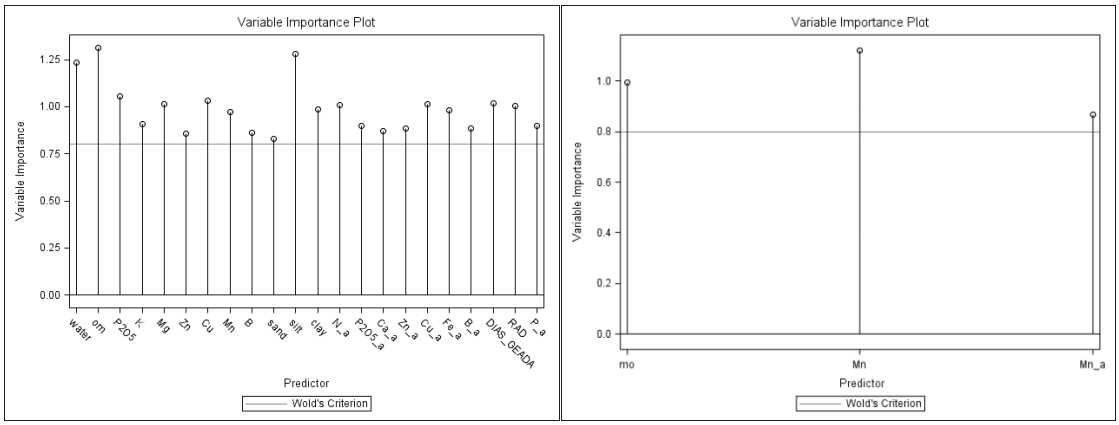


Figure 3. Variance importance plots of the variables in the global FR models (Treatment 1 plots only on the right).

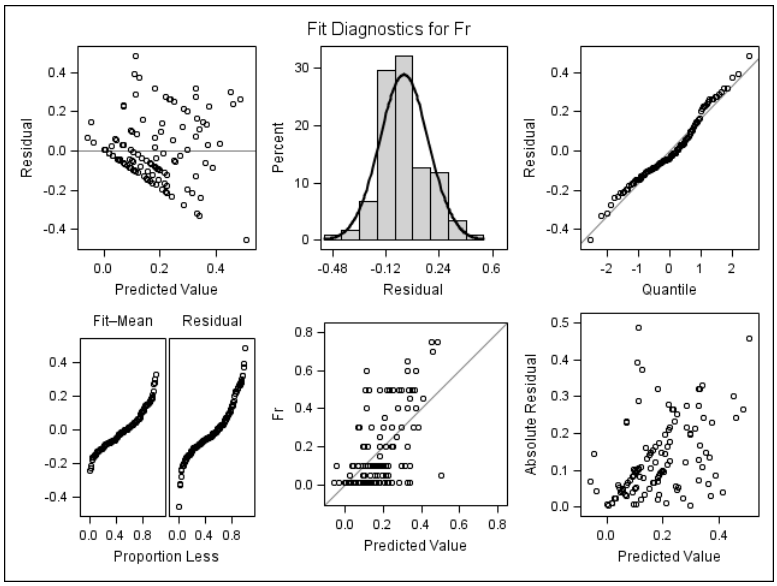


Figure 4. Fit diagnostics for the global FR model.

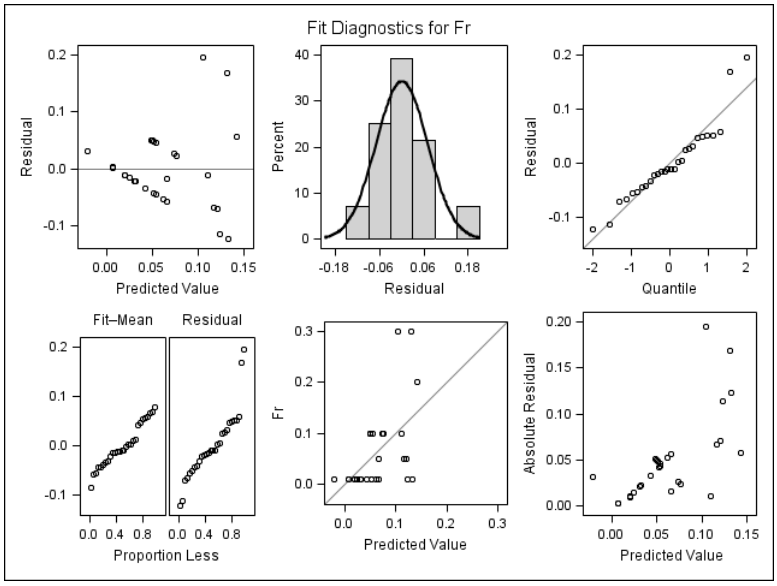


Figure 5. Fit diagnostics for the global FR model with the treatment 1 plots only.

Discussion

Combining PBMs and soil analysis can be a potentially useful tool for fertilization recommendations (e.g. Landsberg *et al.*, 2003; Almeida *et al.*, 2004, 2010), but the limited knowledge of the relationships that exist between FR and soil characteristics still remains a barrier to obtain results that can be transferred to operational use. Some recent efforts have been made to overcome it, but the work of Vega-Nieva *et al.* (2013) was the first to produce a model to predict FR from soil parameters with good results. The idea of this work was to follow the methodologies presented in that work and reach a similar model, but fertilizer responses as a function of soil properties are species-, region- and management- specific (Hart *et al.*, 1986; Gale *et al.*, 1991). The models developed for each of the sites performed well, but the results for the global ones were not so good. However, the results can be considered promising and seen as a starting point for more research. Data from other fertilization trials can provide additional information for improving the model that predicts FR from soil characteristics and to allow its use as a tool in current forest management. Still, one question remains, does the way in which FR is used in the model really reflect the site responses to changes in fertility?

There is an interesting anomaly in relation to tree nutrition and our ability to predict the nutrient requirements of trees and the extent to which soil nutrient status influences, or indeed controls, tree growth rates. The anomaly lies in the fact that there have probably been more experiments aimed at identifying the nutritional requirements of plantation forests than there have been on any other aspect of forest management, with the exception of spacing and thinning trials. Despite this, our ability to predict the nutrient requirements of trees, or the effects of measured amounts of nutrients in soils on tree growth, remains very poor. There has always been a major problem in predicting the nutrient requirements of plantations or the growth rates of natural forests, because growth is not necessarily correlated with the usual measures of nutrient availability in soils, as indicated by the various methods used to determine them. The inability to predict quantitatively the effects of nutrient losses during management or responses to fertilizations results from both the complexity of the soil-plant system and the approach that has been adopted in much of the research in this area. There has been work on the various aspects of soil chemistry aimed at characterizing soils in terms of ability to supply nutrients, but much of the studies are related to total growth observed in the end of a period, not over periods of time.

Smethurst (2000) notes that a major limitation in current soil analysis is the need for extensive field calibration by replication of fertilizer experiments through space and time. He also notes that “concentrations of P in various types of soil extracts have been regressed against relative yield, but these calibrations have a high degree of specificity to type of extract, the soil type, climate, genotype and several other factors.” Some of the difficulties associated with the use

of soil analyses to predict tree nutrients requirements are illustrated by Mendham *et al.* (2002), who made a detailed study of the responses to P fertilization of young eucalyptus plantations in 24 field experiments, in relation to a range of soil P analyses that characterized various pools of soil "available" P. Correlations with the various measures of P were generally low, despite the fact that responses were assessed at a stage (1 year old) when the trees show the largest responses to P fertilization. Clearly, different approaches must be considered.

Managers are concerned with how to achieve the best possible growth rate and yield, so they try to ensure the nutrients requirements of trees. To that end, they require the most possible accurate information about trees nutritional requirements. Although a high number of field experiments regarding optimum fertilizations rates have been conducted, the prediction of nutrients requirements is still somehow uncertain.

Chemical analyses of soils have been used for decades to provide information about soil fertility. Soils may be characterized in terms of the nutrient concentrations per unit mass of soil, or the total mass of N and P and exchangeable cations (K+, Mg++, Ca++ and others) per unit ground area. Such data, which may include information about topsoil and subsoil levels, as well as soils physical characteristics such as bulk density and water-holding characteristics, are conventionally used to determine soil fertility. The problem is to determine the relationship between some measure of the amount of some element in the soil, the uptake by trees and the likely response of trees to addition of given amounts of that element in the form of fertilizer.

Conclusion

The main objectives of this work where to develop a generalized model to predict FR from soil characteristics and to try and find which nutrients are the most limiting for eucalyptus growth and to which the growth shows better responses. Asides from that, some other results were found, namely:

. The differences between the values of available soil water estimated in the field and calculated with the Domingo Santos at al. (2006) expression are very small, which means that using that expression is a good way to estimate this parameter.

. 3-PG is very sensitive to the initial values of biomass, particularly of leaf biomass. The model is also sensible to the available water and an increase in both available water and initial biomass values leads to higher growth rates, but the impact of biomass changes is bigger than the impact of water fluctuations.

. The work developed by Vega-Nieva *et al.* (2013) resulted in slightly different models for FR that included as variables clay (that has an important role in water availability), Ca, Na and K. The model developed by Stape *et al.* (2004) used K, P and cation exchange capacity. The

differences between the results of the models developed in this work and the ones developed for Spain might be related to soil differences or the fact that in this work the fertilization component was also considered. The results don´t allow a definitive answer, but establish a starting point for future studies.

Acknowledgments

This paper is a part of the PhD project of the first author, which is funded by a scholarship (SFRH/ BD/ 46594/ 2008) granted by Fundação para a Ciência e Tecnologia (Portugal).

References

Almeida, A.C., Landsberg, J.J., Sands, P.J. 2004. Parameterisation of 3-PG model for fast-growing Eucalyptus grandis plantations. Forest Ecology and Management. 193: 179–195.

Almeida, A.C., Siggins, A., Batista, T.R., Beadle, C., Fonseca, S., Loos, R. 2010. Mapping the effect of spatial and temporal variation in climate and soils on Eucalyptus plantation production with 3-PG, a process-based growth model, Forest Ecology and Management. 259(9):1730-1740.

António, N., Tomé, M., Tomé, J., Soares, P., Fontes, L. 2007. Effect of tree, stand and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research 37: 895-906.

Domingo Santos, J.M., Fernández de Villarán, R., Corral, E., Rapp, I. 2006. Estimación de la capacidad de retención de agua en el suelo: revisión del parámetro CRA. Invest Agraria: Sistemas y Recursos Forestales 15:14–23.

Fontes, L., Landsberg, J., Tomé, J. Tomé, M., Pacheco, C. A., Soares, P., Araújo, C. 2006. Calibration and testing of a generalized process-based model for use in Portuguese eucalyptus plantations. Canadian Journal of Forest Research. 36: 3209-3221.

Gale, M.R., Grigal, D.F., Harding, R.B. 1991. Soil productivity index: predictions of site quality for white spruce plantations. Soil Science Society of America Journal. 5: 1701–1708.

Hart, S.C., Binkley, D., Campbell, R.G. 1986. Predicting current growth and N þ P fertilization growth response of loblolly pine stands in the Coastal Plain of North Carolina. Soil Science Society of America Journal. 50: 230–233.

Landsberg, J., and Sands, P. 2010. Physiological ecology of forest production. Principles, processes and models. Academic Press. ISBN: 9780123744609.

Landsberg, J.J., Waring, R.H., Coops, N.C., 2003. Performance of the forest productivity model 3-PG applied to a wide range of forest types. Forest Ecology and Management. 172:199–214.

Landsberg, J.J.,Waring, R.H. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning. Forest Ecology and Management. 95: 209–228.

Levy, P.E., Hale, S.E., and Nicoll, B.C. 2004. Biomass expansion factors and root:shoot ratios for coniferous tree species in Great Britain. Forestry, 77(5): 421–430.

Matala, J., Hynynen, J., Miina, J., Ojansuu, R., Peltola, H., Sievänen, R., Väisänen, H., Kellomäki, S. 2003. Comparison of a physiological model and a statistical model for prediction of growth and yield in boreal forests. Ecological Modelling, 161: 95-116.

Mendham, D. S., Smethurst, P. J., Holz, G. K., Menary, R. C., Grove, T. S., Weston, C., Baker, T. 2002. Soil Analyses as Indicators of Phosphorus Response in Young Eucalypt Plantations. Soil Science Society of America Journal 2002. 66: 959–968.

Myers, R. H. 1986. Classical and Modern Regression with Applications. Second edition. Duxbury Classic Series. 357 pp.

Oliveira, T.S., Tomé, M. 2015. Modelling the aboveground stand biomass of planted and coppiced *Eucalyptus globulus* in Portugal. Submitted

Paul, K. I., Polglase, P. J. 2004. Calibration of the RothC model to turnover of soil carbon under eucalypts and pines. Australian Journal of Soil Research. 42: 883-895.

Peng, C.H., Liu, J., Dang, Q.L., Apps, M., Jiang, H., 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. Ecological Modelling. 153: 109–130.

Sands, P. J., Landsberg, J. J. 2002. Parameterisation of 3-PG for plantation grown *Eucalyptus globulus*. Forest Ecology and Management, 163: 273-292.

Sands, P.J. 2003. Process-based models for forest management — integrating determinants of growth into practical management systems. CSIRO Technical Report 126.

SAS Institute Inc. 2009. SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc. 7886 p.

Smethurst, P.J. 2000. Soil solution and other analyses as indicators of nutrient supply: a review. Forest Ecology and Management, 138: 397–411.

Stape, J.L., Ryan, M.G., Binkley, D. 2004. Testing the utility of the 3-PG model for growth of Eucalyptus grandis × urophylla with natural and manipulated supplies of water and nutrients. Forest Ecology and Management. 193: 219–234.

Vega-Nieva, D.J., Tomé, M., Tomé, J., Fontes, L., Soares, P., Ortiz, L., Basurco, F., Rodrígez-Soalleiro, R. 2013. Developing a general method for the estimation of the fertility rating

parameter of the 3-PG model: application in *Eucalyptus globulus* plantations in northwestern Spain. Canadian Journal of Forest Research, 43: 627-636.

Wold, S. 1995. PLS for multivariate linear modelling in QSAR chemometric methods in molecular design. Methods and principles. In Chemistry, medicinal. Edited by H. Van de Watterbeemd. Verlag-Chemie, Weinheim, Germany.

Chapter 5 - Using stand level allometric equations to hybridize process-based and empirical forest growth models. An application to 3-PG calibrated for eucalypt stands in Portugal

Executive summary

A methodology was developed to establish the link between the stand process based model 3-PG and the growth and yield model GLOBULUS. The linkage is accomplished through stand level allometric relationships between basal area and aboveground biomass. The final model is sensitive to silvicultural and climate changes and gives detailed output on stand structure.

Abstract

- Context: Process based models are developed mainly to help understand stand behaviour by simulating the stand growth in terms of the driving physiological processes. They are very useful scientific tools, but tend to require too much detailed input data and give too little diversified outputs to be used often as tools in forest management. In order to make process-based models more operational, there have been efforts towards their simplification and to integrate mechanistic elements with more structurally simple empirical elements. One of the great challenges has been to develop methodologies that allow the combination of the two types of approaches (mechanistic modules combined with empirical functions) in order to take advantage of the strengths of both. The resulting models are often referred to as mixed or hybrid models.

- Aims: The objective of this paper is to develop a methodology to establish the link between a whole stand process based model, 3-PG, and the growth and yield model GLOBULUS. This linkage will result in a final model that is not only sensible to silvicultural and climate changes, but that also gives detailed output on stand structure, such as merchantable volumes and diameter distributions, the type of information that is important on forest management.

- Methods: The linkage between the two models is accomplished through stand level allometric relationships between basal area and aboveground biomass rather than through the relationship between site index and annual production used in some previous model hybridization exercises. The proposed methodology is applied to the 3-PG model calibrated for *Eucalyptus* stands in Portugal, linked to the growth and yield model GLOBULUS. An allometric relationship between dominant height and woody biomass was also developed to allow the use of the GLOBULUS functions that include dominant height as regressor. The output of the hybrid model was also modified by the inclusion of an allometric equation to predict total underbark volume from woody biomass.

- Results: By using the GLOBULUS functions, the hybrid model includes as output not only better predictions of basal area and total underbark volume, but also merchantable

volume to any top diameter, with and without bark, and allows simulating the transition between high forest and coppice and between successive cutting cycles. Stand structure can also be simulated using a diameter distribution model already available.

- Conclusion: The proposed methodology can be applied to other pairs of process-based and empirical forest growth models, allowing that the advantages of process based models are taken into account and that they can be used more often as tools in current forest management.

Keywords: *Eucalyptus globulus*, 3-PG model, GLOBULUS model, hybridization, stand level allometric relationships

Introduction

Prediction of future forest development under alternative management is essential for forest planning. This has traditionally been achieved using the so-called empirical growth and yield models. Empirical models describe forest development with growth functions fitted with large data sets through statistical regression techniques. They are commonly used for forest management and planning, because they supply correct and reliable forest productivity information (Levy *et al.*, 2004), although its use is restricted to the site conditions and forest management practices for which there were growth data available in the fitting data set. Most empirical models make use of the concept of site index (S), an indirect measure of the potential growth at a site assessed by the expected dominant height at a reference age. Implicit in this concept are the assumptions that the dominant height growth of a stand is independent of changes in the environment and is not influenced by stand stocking. When growth conditions are stable, one can use this concept without problems, but when the interest falls on new locations, changing soil or climatic conditions, the growth response is highly non-linear and the relationships can be very difficult to predict. It is known, for example, that the relationship between tree height and diameter varies with altitude in such a way that trees that are grown in higher altitudes tend to be shorter and thicker (Turnbull *et al.*, 1993), resulting in different productivities for the same value of S.

There are methodologies for the empirical growth models to take some of these problems into account, e.g. by developing regional models and/or by expressing some of the parameters of the growth functions as a function of climate and soil variables (e.g. Bravo-Oviedo *et al.*, 2008; Nunes *et al.*, 2011) but these methodologies are not able to cover all the situations that may occur. The use of S as driving variable in growth models is not the best approach, not only under a climate change scenario, but also because simulation of silvicultural techniques that change site productivity, such as irrigation and or fertilization, may have impact on the shape of the dominant height growth curve and therefore on S estimates.

The use of empirical models is compromised when there is no information about previous conditions, such as new plantations or when the site conditions change due to fertilization or climate changes, for example. Additionally, the changing environment under which the forests are developing nowadays, namely climate change and management intensification, makes the use of empirical models less appropriate. The use of process-based models to make projections of forest development is therefore gaining relevance. Such models, integrating the main physical, biogeochemical and physiological processes involved on forest growth and development, give a mechanistic description of the interactions between the living plants and their environment and are able to assess the energy balance and the cycling of water, carbon

and nutrients within a given ecosystem. Process based models simulate the stand growth pattern in terms of the physiological processes that determine growth and have been developed mainly to help understand the stand behaviour from the plant-soil and water-carbon-nutrient points of view, being in this way useful for long term predictions, especially considering silvicultural and climate changes. They are mathematical representations of biological systems that incorporate knowledge about physiological and ecological mechanisms into prediction algorithms (Landsberg, 1986) being a structure that synthesizes the available knowledge. They are also scientific tools that may help to formalize and test hypothesis. However, process based models have not been used very often as tools in forest management because managers consider that they encompass many uncertainties and require difficult to obtain parameters which imply that their projections may not be as reliable as the ones from empirical models (Mohren and Burkhart, 1994). Besides, the idea persists that the statistical conventional approach to growth and productivity modelling, based on large data sets, is superior (Mäkelä *et al.*, 2000). Process based models tend to be more complex and they require a lot more of input data than empirical models, one of the reasons why the last are not more commonly used for operational forest growth predictions.

In order to make process-based models more operational, there have been efforts towards their simplification and to integrate mechanistic elements with more structurally simple empirical elements. One of the great challenges has been to develop methodologies that allow the combination of the two types of approaches (mechanistic modules combined with empirical functions) in order to take advantage of the strengths of both. The resulting models are often referred to as mixed or hybrid models. Hybridization of process based and empirical models is seen as an efficient way to provide forest managers with forest growth information under a changing environment, using inputs and providing outputs compatible with operational forest management while applying relevant physiological knowledge. In the last years, many were the authors that tried this approach, known as hybridization of process based and empirical models (Almeida *et al.*, 2003; Baldwin *et al.*, 2001; Bataglia *et al.*, 1999; Luxmoore *et al.*, 2002; Matala *et al.*, 2005; Matala *et al.*, 2006; Peng *et al.* 2002; Robinson and Ek, 2003). The first hybrid models based the linkage between a process based and an empirical model, most of the times, by adjusting the site index of the empirical model using biomass production estimated by the process-based model. For instance, Baldwin *et al.* (2001) linked the empirical individual tree model PTAEDA2 (Burkhart *et al.*, 1987) with the process-based model MAESTRO (Wang and Jarvis, 1990). PTAEDA2 predicts the characteristics of trees grown at a given stand density, on a given site, for a given length of time, its outputs being used as direct inputs into MAESTRO which assesses the expected impact of environmental changes. The results from MAESTRO are then used to predict the change in site index as a function of

net photosynthesis, age, and stand density. This information is fed back into PTAEDA2 to update future predictions by modifying site index. A similar methodology is found in Battaglia *et al.* (1999), where the relationship between S and the mean annual increment is used: the process based model predicts the biomass growth from which the mean annual increment in volume (i_V) is estimated and S is then predicted as a function of i_V. The site index value is used as an input for the empirical model, which is going to provide the usual and diversified outputs expected for this kind of models. Tomé *et al.* (2004), recognizing that S is not appropriate to establish the linkage between the models, proposed the use of the allometric relationship between stand basal area and aboveground biomass. Others, like TRIPLEX (Peng *et al.*, 2002) and Forest 5 (Robinson and Ek, 2003) are hybrid models that take advantage of components of existing models, combining them together and creating new models.

According to Mäkelä (2009), followed by Kimmins *et al.* (2010), the hybrid models may be classified in 3 categories: i) Hybridized empirical models or models that utilize process elements as submodels/variables to provide additional inputs to empirical growth models. The inputs are often expressed as modifiers of the growth functions developed statistically, including, e.g., intercepted radiation, gross or net photosynthetic production, soil water status, or summary functions derived from full process model simulations. Examples of this methodology are CanSPBL(water) (Pinjuv *et al.*, 2006) or MELAFinnFor (Matala *et al.*, 2005; 2006); ii) Hybridized process based models that derive potential growth from empirical data or functions, modified using a complex system of physiologically-based functions or submodels. This group includes models of the gap model family (see e.g. Bugmann, 2000, for a revision on gap models), and ecosystem models such as FORECAST (Kimmins *et al.*, 1999); iii) Process-driven hybrid models where growth is derived explicitly from carbon acquisition and allocation. Empirical elements relate, e.g., to the allocation of carbon between and within trees, but also to the process of parameterization and calibration of the model as a whole. Examples are: 3-PG (Landsberg and Waring, 1997), PipeStem, PipeQual (Mäkelä, 1997, 2002; Mäkelä and Mäkinen, 2003), TRIPLEX (Peng *et al.*, 2002). The research described here focus process-driven hybrid models. In fact, contrary to the other two categories, these models predict growth in a process-based way, even if combined with empirical prediction of other state variables. Ultimately all process based models rest on some empirical support, therefore they may be considered as part of this category of models.

The objective of the research described here was to test the methodology proposed by Tomé *et al.* (2004) to develop a hybrid model by using a whole stand process-based model (in the sense of Mäkelä's categories it may be called a process-driven hybrid model) linked to an empirical model without using site index to establish the link between the two models, originating a model that is sensitive to climate change and silviculture intensification and

produces as output all the variables usually available from empirical models. The proposed methodology is applied to hybridize the process based model 3-PG (Landsberg and Waring, 1997) with an empirical growth and yield model, the GLOBULUS model (Tomé *et al.*, 2006b) so that the diversified output of this model can be available to managers. The proposed methodology can be applied to other pairs of process-based and empirical model.

Material and methods

II.1. The hybridization procedure

A forest growth model is a simplified representation of the forest through the values of a set of variables, the state variables. State variables may describe the forest and/or the trees (e.g. basal area, standing volume, diameter of the trees) or some compartments or pools of the ecosystem (soil water, soil nutrients). Generally, a forest growth model includes a growth (dynamic) module and a prediction module. The growth module predicts the evolution on time of a sub-set of the state variables – the principal variables – while the prediction module includes a set of equations that predict, for a certain point in time, the values of the remaining state variables – the derived variables – using the values of the principal variables and, eventually, other derived variables as input. The growth module is the core of the model and the distinction between empirical and process-based models lies precisely in this module. In an empirical model, the growth module includes a set of equations for the projection of the principal variables over time, using a pre-defined time step, usually the year. In a process based model, the growth module includes the modelling of at least some of the processes involved in the evolution of the principal variables, reason why they usually use smaller time steps such as the day or the month. The principal variables are usually different in the two model types. Process based models usually predict biomass production and its allocation to the different plant components: roots, leaves, stem, branches and fruits. The principal variables are the tree biomass pools and soil water and nutrients pools. Management oriented stand variables, such as basal area, volume (under and over bark), merchantable volumes or tree level information need to be predicted from the biomass estimates. Empirical models, on its turn, encompass a set of growth functions that directly simulate growth of the management oriented variables taking into account site quality, usually through site index and stocking. Environment influence is taken into account through external variables – the control variables – also different in the two types of models, with much more detail and impacting on biomass production and allocation in process based models and just adjusting the values of the parameters in empirical models. Additionally, both model types include a silviculture module

that comprises a series of algorithms to simulate silvicultural treatments such as thinning or pruning and, eventually, models to simulate the occurrence of hazards.

As mentioned in the introduction, the hybridization is a combination of a process based model that can reflect environmental changes and an empirical model that gives detailed output for operational purposes. In this way the new model will have the strengths of both models. The methodology used in this paper proposes the development of a hybrid model composed of:

1. The growth module from a process-based model – eventually a process-driven hybrid model in the nomenclature of Mäkelä (2009) – that estimates the evolution of the biomass pools (principal variables) over time;
2. A set of equations – the link functions – specifically developed to establish the link between the principal variables of the process based model and the variables needed as input in the prediction equations from the empirical model;
3. The equations of the empirical growth and yield model to predict the dependent variables not included in 1. or 2.

According to this structure, the hybrid model will include the stand biomass per plant component (for instance root, leaves and woody parts) as principal variables predicted with the process-based model. Some of the management oriented variables, for instance stand basal area, will be selected as link functions and modelled in a way to establish the link between the process-based and the empirical models. The remaining management oriented variables, such as volume (under and over bark), merchantable volumes or tree level information (diameter distributions), are predicted using the equations from some existing empirical model (that may be specifically developed for the purpose, if not previously available). Silviculture algorithms used in the empirical models may also be “imported” for the hybrid model.

The proposed methodology is applied to develop a hybrid model for eucalypt plantations that combines the growth module of the 3-PG model (Landsberg and Waring, 1997) with a set of prediction equations taken from the GLOBULUS empirical model (Tomé *et al.*, 2006b). A brief description of both models is given in sections 2.3 and 2.4.

II.2. The 3-PG model

A process based model that has revealed itself as interesting, easily adaptable to different species and with good results in eucalypt stands is the 3-PG (Physiological Principles in Predicting Growth) (Landsberg and Waring, 1997; Sands and Landsberg, 2002). It is a simple, physiologically based model developed by Landsberg and Waring (1997) with the intention of bridging the gap between the simpler empirical models and the more complex physiological

based ones, giving stand level information with interest for the forest management (volume under bark and the respective annual increment, stand basal area and quadratic mean diameter). It is not too demanding in terms of input data, it has the ability to reflect the effect of changes in the environment, is well documented and the code is available. The research group in which the present research was conducted has been working on this model and it has already been calibrated for *Eucalyptus* plantations in Portugal (Fontes *et al.*, 2006). Table 1 summarizes the state variables that make part of the growth and prediction modules of the 3-PG model as well as the respective control variables. The model requires a few parameters, climatic data inputs, basic knowledge about the local soil water-holding capacity and an indication of soil fertility. According to his authors it is structured in five sub-models that describe 1) the assimilation of carbohydrates, 2) distribution of biomass between foliage, roots and stems, 3) soil water balance, 4) determination of stem number and 5) conversion of biomass values into variables of interest to forest managers (Sands and Landsberg, 2002). The model works at a monthly time step and sub-models 1) to 3) form the growth module with the following principal variables: gross primary production (GPP), net primary productivity (NPP), leaf area index (LAI), standing biomass (partitioned into foliage, woody biomass and roots) and soil water pool. 3-PG may be included under the APAR (absorbed photosynthetically active radiation) models-type (Battaglia and Sands, 1999) since the production of biomass is based on light interception via Beer's law with light use efficiency modified by environment and tree age and a constant ratio of NPP to GPP. Allocation to roots depends on growing conditions and aboveground biomass partitioning is based on the allometric relationships between tree biomass components and total above ground biomass, being allocation to leaves affected by tree size. Soil water balance is estimated through a single soil layer model with evapotranspiration determined from Penman-Monteith equation.

Table 1. State and control variables included in the 3-PG model of Landsberg and Waring (1997)

Control variables	Principal variables (growth module)	Derived variables (prediction module)
Environmental Soil: texture, fertility ratio (0-1), maximum available soil water Site: latitude **Climate** Mean temperature, solar radiation, vapor pressure deficit, precipitation, days of frost **Management** Initial stand density, thinning, fertilization, irrigation	Biomass production: Root biomass (W_r) Leaf biomass (W_l) Woody biomass (W_{wy}) Soil water balance: Available soil water (Asw)	Diameter of the tree with average W_{wy} (d_{wy}) Stand basal area (G) Underbark volume without stump (V_{u_st}) Net primary production (NPP) Evapotranspiration

Sub-model 4) takes care of mortality using the 3/2 power law, with the possibility of predicting additional mortality with an empirically fitted mortality function. The prediction module coincides with sub-model 5) and includes the following derived variables: diameter of the tree with average woody biomass (d_{wy}), stand basal area (G), volume under bark without stump (V_{u_st}) and the respective mean annual increment (i_{Vu_st}).

V_{u_st} prediction is based on the prediction of woody biomass (wood + bark + branches), one of the outputs of the 3-PG growth module. Woody biomass (W_{wy}) is converted into wood biomass through the estimation of the ratio between branches+bark and the woody biomass (p_b) that is then divided by wood density (ρ) to provide the estimate of V_{u_st}:

(1) $$V_{u\_st} = \frac{W_{wy}\,(1-p_b)}{\rho}$$

Both p_b and ρ are modelled as a function of age.

G prediction is achieved through the estimation of the average tree diameter from woody biomass and stem number. First, the woody biomass of the mean tree is predicted by division of W_{wy} by N. The diameter of the tree with average woody biomass (d_{wy}) is then estimated by an "inversion" of the allometric equation for tree woody biomass prediction. This diameter is used to compute basal area of the mean tree (g_{med}) which is then multiplied by N to provide the G estimate:

(2) $$\overline{w_{wy}} = \frac{W_{wy}}{N}$$

(3) $$\overline{w_{wy}} = k\, d_{wy}^{a} \quad \Rightarrow \quad d_{wy} = \left(\frac{\overline{W_{wy}}}{k}\right)^{1/a}$$

(4) $$g_{med} = \frac{\pi}{4}\left(\frac{d_{wy}}{100}\right)^{2}$$

(5) $$G = N\, g_{med}$$

where k and a are the scale parameter and the allometric constant of the allometric relationship between w_{wy} and d. This method has several problems: stand variables based on the mean tree tend to give biased estimations; when a regression of y is fitted over a regressor x, the estimations of x obtained by "inverting" the regression are usually biased; small variations on the values of the scale parameter (k) and of the allometric constant (a) of the allometric equation that provides the estimates of d_{wy} lead to large variations in the values of G and stand volume due to the non-linearity of the equations (Landsberg *et al.* 2003). Additionally, when using data from the Portuguese data-set mentioned above, the relationship between w_{wy} and tree d is not as clear as shown in Sands and Landsberg (2002) (Figure 1).

For detailed descriptions of the model the readers are remitted to Landsberg and Waring (1997) and Sands and Landsberg (2002).

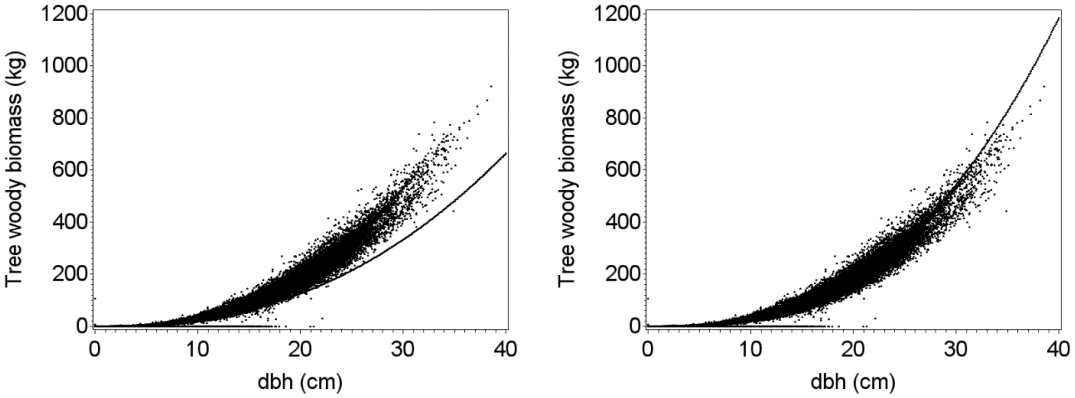


Figure 1. Tree woody biomass and diameter at breast height (dbh) relationship. The line is the fitted relationship used in 3-PG (Sands e Landsberg, 2002) with k=0.095 and a=2.4 on the left and the fitted relationship used in the Portuguese calibration of 3-PG (Fontes *et al.* 2006, 2002) with k=0.056 and a=2.7 on the right.

II.3. The GLOBULUS model

GLOBULUS (Tomé *et al.* 2001; Tomé *et al.* 2006b) is an empirical stand level growth and yield model developed for Portuguese *Eucalyptus globulus* pure even-aged stands. It integrates all the available information on eucalypt growth and yield in Portugal and represents the combined efforts between industry and universities, which have been involved in several cooperative research projects over the past decades. The model is regionalized by adjusting the parameters of the equations depending on each particular stand. Initially, the country was classified in homogeneous climatic zones and, based on these regions, the model parameters were adjusted according to the stand location (Tomé *et al.*, 2001). In the last version of the model, GLOBULUS 3.0 (Tomé *et al.* 2006b), the idea was to associate to each stand a group of site-related variables, such as precipitation, temperature, altitude, among others, that are used to allow some of the parameters to be stand-specific. The main differences between GLOBULUS 3.0 and the previous versions are (Tomé *et al.*, 2001): i) it expands the parameters of the growth functions as linear functions of climatic and other site-related variables; ii) it simulates stand growth for coppiced stands prior to the thinning of the shoots, that usually occurs by the age of 3 years, allowing the possibility to simulate the transition between cutting cycles and the stand behaviour prior to the thinning operation; and iii) it includes improved volume and biomass equations. It has three modules: initialization, growth and prediction. The

growth module projects the principal variables over time, while the prediction module predicts the values of the derived variables at each point in time. The initialization module predicts each principal variable as function of control (external) variables and is essential to estimate the values of the principal variables in stands that have not been measured yet and also in new stands, either planted or coppiced. The growth module consists of a set of compatible equations for the principal variables that are function of the values of the variables at the beginning of the growth period and of the control variables. The growth functions are all formulated as first order non-linear difference equations. Table 2 lists all the variables that are currently included in the model. The equations can be seen in Tomé *et al.* (2006b).

Table 2. State and control variables currently included in the GLOBULUS 3.0 model

Control variables	Principal variables (growth module)	Derived variables (prediction module)
Environmental: Site Index (base age 10) Elevation **Climate**: Precipitation and precipitation days, mean annual temperature, frost days **Management**: Initial stand density (1st rotation) Stand density after thinning (coppice) Number of rotations Harvesting age in each rotation Rotation	Dominant height (hdom) Number of trees per hectare (N) Stand basal area (G) Total volume (with stump) under bark (V_u), bark volume (V_b) and stump volume under bark (V_{st})	Merchantable volumes (under bark) (V_{mudi}) Total biomass (aboveground) (W_a) Biomass per plant component: branches, leaves, wood, bark (W_{br}, W_l , W_w , W_b) Carbon stock per plant component Number of living stools Shoot number before and after thinning

II.4. Data

Data for this study came from the large database available in Portugal for eucalypt stands. It includes data from permanent plots, experimental trials and continuous forest inventory of the pulp companies. The available data set, after editing for deletion of abnormal points, was analysed and validated and stand level variables estimated from the tree level information. Individual tree heights, measured just in sample trees, were estimated with the height-diameter curve from Tomé *et al.* (2007a). Different stand volumes – total volume without stump over and under bark, bark volume, merchantable volumes to top diameters between 10 and 5 cm – were computed using the system of volume equations from Tomé *et al.* (2007b). Aboveground

biomass and biomass per tree component (stem wood, stem bark, branches and leaves) were estimated using the system of biomass equations developed by António *et al.* (2007). Crown length, used as regressor in the leaf and branches biomass equations, was estimated with the crown ration and crown length system of equations developed by Oliveira and Tomé (submitted).

The data set used contains information on nearly 7,000 growth periods (t_1 and t_2 observations for the development of growth equations) covering a wide range of ages, stand densities and site index (Table 3).

Table 3 – Distribution of the data by age, site index and stand density

S, base age 10 (m)	Stand density (trees/ha)	Age				Total
		t ≤ 4	4 < t ≤ 8	8 < t ≤ 12	t > 12	
S ≤ 16	N < 1111	5	95	403	87	590
	1111 ≤ N < 1667	2	61	143	47	253
	N > 1667		4	17	43	64
16 < S ≤ 20	N < 1111	21	376	942	209	1548
	1111 ≤ N < 1667	22	294	541	103	960
	N > 1667	27	60	119	58	264
20 < S ≤ 24	N < 1111	28	318	665	181	1192
	1111 ≤ N < 1667	34	247	495	107	883
	N > 1667	49	93	156	64	362
S ≥ 24	N < 1111	18	82	149	70	319
	1111 ≤ N < 1667	18	105	151	62	336
	N > 1667	19	63	46	35	163
Total		243	1798	3827	1066	6934

II.5. Details on the hybridization / improvement of the 3-PG prediction module

In 3-PG, total biomass (net primary production) is the principal variable directly estimated in the growth module. The allocation module distributes the biomass produced into the plant components: roots, leaves and woody plant parts. In the exercise described here no attempt was made to improve or modify the allocation module. The focus was just the module that provides information for managers (the prediction module). The goal is to improve this module by developing link functions for G, V_{u_st} and hdom that predict these variables from the biomass estimates provided by 3-PG, allowing the use of the equations of GLOBULUS for other

management oriented variables and therefore providing estimates for all the variables provided by the GLOBULUS model. In the cases of G and V_{u_st}, the performance of the link-functions must be compared with the performance of the original equations used in 3-PG (see section 2. The 3-PG model).

II.5.1. Developing link functions for G, V_{u_st} and hdom

Due to the reasons stated previously, the linkage between the two models was not accomplished by an adjustment of S, but by exploring the allometric relationships between each one of the variables selected to establish the link between the two models and some biomass component.

Stand basal area (G) is the most important principal variable in the GLOBULUS model, used as regressor in the prediction of most of the state variables, either principal or derived variables. For this reason it was selected as one of the link variables. The analysis of the structure of the GLOBULUS model shows that dominant height (hdom) is also used as regressor for most of the sub-models. Therefore, a second link-function had to be developed for dominant height (hdom-link function).

Stand volume (with or without stump and with or without bark) can be estimated using the volume equations from the GLOBULUS 3.0 model that require hdom and G, predicted with the link functions, as regressors. However these predictions already have an associated error, which will be amplified in the final volume estimate. Under bark volume without stump is one of the most important variables for operational purposes. By this reason, and also because under bark volume without stump is one of the outputs of 3-PG, it was decided to develop also a link function for under bark volume without stump (V_{u_st}).

The three link functions were developed using a similar methodology and basal area (G) is used here to explain the methodology. The G-link function is based on the allometric relationship between G and some biomass component W (equation 6a) that can be expressed as a difference equation by establishing the ratio between G at two points in time (t1 and t2):

(6a) $G_t = k_t W_t^{a_t}$ (allometric relationship – initialization)

(6b) $G_{t2} = G_{t1} \frac{W_{t2}^{a_{t2}}}{W_{t1}^{a_{t1}}}$ (difference equation – projection)

where k_{ti} and a_{ti} are the scale parameter and the constant of the allometric relationship (the subscript t or ti indicates the value of a at age t or ti) and t1 and t2 are the beginning and end of the target growth period. Different biomass components – total aboveground, woody, leaves, crown and stem biomass – were tested and the one that lead to a more precise and less biased

model was selected (see section 3 of Results). The advantage of expressing the allometric parameters k_t or a_t as a linear function of stand variables, namely stand density, stand age and a dummy variable for coppice, was examined:

(7a) $$k_t = k_0 + \sum ak_i \, X_{it}$$

(7b) $$a_t = a_0 + \sum a_i \, X_{it}$$

Where X_{it} is variable X_i at time t and ki and ai are model parameters.

When using the model, the biomass growth for the period [t1, t2] is predicted by 3-PG and the current biomass (W_{t2}) is obtained by adding the estimated growth to the previous biomass (W_{t1}). Basal area for the current period (G_{t2}) is then predicted using the allometric relationship between basal area and biomass expressed as a difference equation (equation 6b). The link function can also be used in the allometric form (equation 6a), when there is no information on basal area at the start of the growth period.

II.5.2. Selection of the allometric models for each one of the variables G, hdom and Vu_st (link-functions)

On a first stage, a test was made with the different biomass components (aboveground, woody, stem, wood, leaves, crown and branches) to find out which one related the best with each of the variables to be modelled (G, hdom and V_{u_st}). This first analysis was made using the PROC NLIN procedure (SAS Institute Inc., 2009b). Then, the allometric constant a of each model was expanded as a function of different combinations of stand variables in order to select the variables to be included as regressors in the expressions of the a parameter. The candidate regressors were N, t and a dummy variable for coppice (0, if a seedling stand; 1, if stand regenerated by coppice). In most equations, a scaling factor (N/1000 or t/10, for example) was used so that the coefficient estimates where, whenever possible, on the same scale (Schabenberger and Pierce, 2002). A similar analysis was made for the scale parameter k.

The comparison of the alternative models was based on the values of the residual sum of squares (SSR) and R^2.

This analysis allowed the selection of the biomass component to be used in each one of the link functions and also to assess the importance of expressing the allometric parameters (a and k) as a function of each one of the stand variables considered, leading to the formulation of a preliminary model for each link function. The opportunity of expressing both parameters of the allometric relationship as a function of stand variables was analysed at this stage. Multiple allometric relationships, with more than one biomass component, were also tested. This procedure led to the selection of a set of models for further analysis.

II.5.3. Model evaluation and comparison with the 3-PG model

Model evaluation begins with the theoretical aspects of model building, such as the assurance that the regression assumptions have been checked, and in the analysis of the logic of its structure and of the biological aspects. Besides from that, model performance, namely bias and precision should be evaluated with data different from those used in the model fitting. When there is no additional data set for model evaluation, resampling procedures can be used (Burkhart and Tomé, 2012). The evaluation of the models was made considering i) the verification of regression assumptions; ii) the analysis of the signs of the parameters associated with the allometric parameters, iii) assessment of model performance using a resampling procedure.

The regression assumptions, normality, independence and homoscedasticity of the model errors, were checked on the models selected. The heteroscedasticity of the errors was assessed through the analysis of the plots of the studentized residuals versus the predicted values. In case of nonhomogeneous variance of the errors, weighted regression was used. The residuals were weighted, in an iterative process, with the inverse of the standard error of the estimate. The normality of the residuals was analysed with the help of the normal QQ plots and corrected, when necessary, with robust regression using the Huber function (Myers, 1986).

The presence of auto-correlation among the model errors (due to the existence of repeated measurements on the same plot) was analysed using the ODS tool of SAS (SAS Institute Inc., 2009c). The plots of the autocorrelation function (acf) and partial autocorrelation function (pacf) were analysed in order to decide which structure best fitted the data: AR, MA or ARMA and its order.

The sign expected for each one of the parameters (positive/negative) was found through the graphical analysis of the allometric relationships in selected subsets of the data set in which it was possible to isolate the impact of one stand variable (stand density, age or the dummy variable coppice). The signs of the estimates were then compared with the respective expected sign.

In order to assess model performance, the data set was sampled 30 times by randomly selecting 20% of the plots, thus obtaining two data sets, the larger (80%) for model fitting and the reminder (20%) to evaluate model performance. The model under evaluation was then fitted 30 times with the fitting data set and prediction values computed with the correspondent evaluation data set. The following statistics based on the prediction residuals ($r_p = y_i - \hat{y}_i$) were used to evaluate model bias and precision:

1. mean value of the r_p residuals, $\overline{r_p}$ (model bias);

2. mean of the absolute value of the r_p residuals, $\left|\overline{r_p}\right|$ (model precision).

Visual/graphical inspection of observed values against model predictions was undertaken and the bias on the observed-to-predicted fit was tested by examining if the values where or not very distant from the 1:1 reference line. Plots of the residuals versus stand variables (stand density, stand age and site index) were scrutinized in order to detect possible bias in relation to stand variables. Particular attention was given to the evaluation of the performance of the link functions for G and V_{u_st} in comparison with the original 3-PG equations.

This analysis allowed, in some cases, for the simplification of the models selected during the first stage as some of the coefficients associated to stand variables previously selected for the linear expansion of the model parameters did not significantly differ from zero for all the re-sampling data sets.

In this evaluation stage, the multiple allometric model for dominant height revealed some problems. For values over 25m, there was a clear overestimation of the value of hdom, even if the values of the evaluation statistics indicated that the results were better with the multiple allometric model when compared with the simple allometric model. The analysis of the residual plots showed that there was an evident relationship with stand age, so this variable was included in the model which led to a reduction in the bias for values greater than 25m but still to unacceptable values. A more thorough analysis showed that the inclusion of the autocorrelation structure, not taken into account during the prediction, was responsible for this tendency and so the autocorrelation structure was not considered in the adjustment of the hdom equation.

II.5.4. Simultaneous fitting of the selected models

After the selection and evaluation of the allometric models for each one of the link-functions (G, V_{u_st} and hdom), a simultaneous fitting of the system composed of the 3 allometric equations plus the corresponding difference equations (see section 5.1 above) took place. The simultaneous fitting is justified to ensure the compatibility between each pair of prediction and projection equations and also because the independent variables appear in more than one function and the dependent variables of some functions appear as regressors in other functions. Since the set of equations has contemporaneous cross-equation error correlation (also known as nonlinear seemingly unrelated regression system), the fitting was made using PROC MODEL (SAS Institute Inc., 2009a) and the ITSUR option, which provides an iterative estimation of the cross-equation covariance matrix. The initial parameters and weight values

for the simultaneous fitting were the ones obtained in the preliminary fitting of the 2 equations for each one of the link-functions.

Note that the number of data used for the simultaneous fitting was smaller than the one used in the selection of the allometric models because the fitting of difference equations requires the use of lagged data which implies the loss of one observation per plot.

The performance of the final system of link functions was assessed with the PRESS residuals with a procedure similar to the one used to evaluate the individual allometric models.

Results

III.1. Selection of the allometric models

Table A.1. in the appendices shows the residual sum of squares (SSR) for alternative link functions for G, V_{u_st} and hdom. The different alternatives refer either to the biomass component used as regressor and/or to the linear function used to expand the allometric constant a. All the three stand variables show a very good relationship with W_a, W_{wy}, W_s and W_w (Table A.1.). Stand basal area also shows a very good relationship with the biomass of crown variables, which is not true for stem volume or dominant height. Considering just the values of the statistics, one would be tented to select the W_{br} as the regressor for the G allometric relationship, but 3-PG only predicts directly W_{wy}, W_l and W_r, so these were the variables considered as predictors on the equations. It was decided to model G based on W_a, since $W_a = W_{wy} + W_l$. The best relationships for V_{u_st} and hdom are with W_{wy}, W_s and W_w, with statistics values very similar between them, so woody biomass was the chosen biomass component for both equations.

Analyzing the impact of expanding the a parameter as a linear function of stand variables, N appears to be the most important variable when the allometric relationship is based on total aboveground, woody or stem biomass, reducing considerably the value of the SSR when used as the single variable in the expansion of the a parameter. The other possible stand variables, t and coppice, even when associated to coefficients statistically different from zero, do not imply large reductions in SSR. On the contrary, when the allometric relationship is based on leaves, crown or branches biomass, stand age is the variable that, when included in the expansion of the a parameter, reduces SSR the most. The results for the k parameter are similar but the expansion of the a parameter should be preferred as this parameter explicitly appears in the difference forms of the allometric relationships

Based on the results in Table A.1., it was decided to model G from W_a, expanding the a parameter according to stand density (N) and V_{u_st} and hdom from W_{wy}, expanding also the a

parameter according to stand density. The parameter associated with N is positive in the G and V_{u_st} models but negative in the hdom model.

Expanding also the k parameter as a function of stand variables led to a reduction in the SSR, therefore the final set of models were allowed to include N in both allometric parameters (Table 4). When testing the use of multiple allometric models the conclusion was that using leaf biomass as a second variable in the allometric relationships largely reduced the SSR. These multiple allometric models are also shown in Table 4. When used to predict stand variables in 3-PG, these models will include the prediction errors of two biomass components which may increase the final prediction error. In other applications of the allometric equations, for instance in the estimation of stand variables from remotely sensed data, the full model may be preferable. It has been decided to include in this paper two systems of equations, one with the simple allometric equations and a second one with the multiple allometric equations, leaving to the users the possibility to choose among them, depending on their objective.

Table 4 – Residual sum of squares (SSR) and values of the statistics used to evaluate bias and precision for the link functions for *G*, V_{u_st} and *hdom* considering simple and multiple allometric models

		Allometric model					
	3-PG	simple a	simple k	simple a, k	multiple a	multiple k	multiple a, k
		Basal area (G, m^2 ha^{-1})					
SSR		7505,8	7194,1	6946,3	1328,5	1324,551	1324,4
Bias - $\bar{r_p}$	– 0,1460	0,0106	0,0193	0,0275	0,0472	0,0491	0,0480
Precision - $\lvert\bar{r_p}\rvert$	1,1625	0,4749	0,4641	0,4561	0,1950	0,1952	0,1948
		Volume (V, m^3 ha^{-1})					
SSR		10890,1	10019,1	9540,9	9858,1	9328,7	9054,4
Bias - $\bar{r_p}$	– 1,3050	– 0,1309	– 0,1221	– 0,1136	– 0,1148	– 0,1098	– 0,1052
Precision - $\lvert\bar{r_p}\rvert$	1,8401	0,5421	0,5291	0,5200	0,5194	0,5101	0,5050
		Dominant height (hdom, m)					
SSR		20501,4	19791,9	19840,5	4198,651	4232,2	5129,3
Bias - $\bar{r_p}$		0,0571	0,0462	0,0423	– 0,0759	– 0,0779	– 0,0774
Precision - $\lvert\bar{r_p}\rvert$		0,8134	0,7977	0,7968	0,3673	0,3678	0,3682

III.2. Evaluation of the selected allometric models and 3-PG comparison

No relevant heteroscedasticity was found from the analysis of the plots of studentized residuals over the predicted values. The analysis of the QQ plots showed some evidence of non-normal distribution of the model errors, so a correction was made with robust estimation with the Huber´s function (Myers, 1986). The analysis of the autocorrelation plots showed that the autocorrelation function (acf) rapidly decreases to zero and that the partial autocorrelation function (pacf) abruptly cuts after lag 1, the typical patterns for a first-order autoregressive model – AR (1) – for the autocorrelation structure of the errors. The value close to zero of the plotted pacf at lag 1 means that the AR(1) model removes all the autocorrelation. The AR process expresses the series in terms of past observations and the current disturbances (random error), so it makes sense that this is the error structure that best fits the data. All equations were fitted considering this error structure, except for the hdom multiple allometric function, as explained in the methods section.

Having checked all the regression assumptions, the analysis of the signs of the parameters was undertaken. This was not a simple task, since the equations have combinations of variables with difficult to interpret interactions. For example, the negative signal on the parameter associated with density in the volume equation doesn´t simply means that there is an inverse relationship between N and V_{u_st}, since other variables are present in the equation.

N shows a positive relationship with G and a negative one with t, which makes sense, more trees mean more basal area and younger trees mean less basal area but when the biomass component used in the equation is leaf or crown biomass, the relationships are the opposite ones. The same thing happens for the dummy variable coppice, is has a negative relationship with G, except when Wl or Wc are used.

Volume presents a positive relationship with N and the coppice variable and a negative one with t, having the inverse relationships when Wl or Wc are used. The dominant height maintains a negative relationship with N and a positive one with t and coppice, regardless of the biomass component used.

Figure A.1 on the appendices presents the result of the evaluation of the hdom prediction and projection equations. The simple allometric equation presents good results and no apparent bias is revealed on the plots, which is consistent with the results presented in Table 4. As for the multiple allometric model, the values are much closer to the 1:1 reference line in the observed-to-predicted plot, which means that this equation predictions are more accurate than those obtained with the simple one. As for the residuals plots, there is a visible improvement in a global scale, but the early detected relationship with stand age is still visible. For ages over 20 years, there is some tendency to over-estimate the hdom values and it might be best to use

the simple allometric equation that shows a best behaviour and an almost zero trend in the plot of the residuals versus age. As for the projection equations, both present very good results and the bias values are largely reduced. The multiple allometric model presents an almost 1:1 relationship in the observed-to-predicted plot and the residuals values are also very close to 0, which means that this is a very precise model.

Figures A.2 and A.3 on the appendices show the results obtained when evaluating the performance of the allometric models in comparison with the equations of the "Information to managers" module of the 3-PG model. The models proposed in this paper give predictions that are much better than those given by the original 3-PG. The 3-PG model clearly underestimates basal area for values higher than 30 m^2 ha^{-1}, which is not the case for the proposed model. The multiple allometric function, that includes WI as additional predictor, tends to reduce the prediction errors, but for values over 40 m^2 ha^{-1} it has some tendency to underestimate G values. Since the amount of leaves increases until a certain age and then the leaf biomass stabilizes, while the basal area continues to increase, this behaviour is understandable. The analysis of the graphics over age, stand density and site index allows the association of the underestimation of the multiple function with older and better site index stands.

The projection equations do not present this behaviour, giving accurate predictions with low values of errors.

The performance of the two models (3-PG and the proposed model) for volume prediction is similar, although the proposed model is slightly more precise and less biased. All the models tend to overestimate volumes over 400m^3 ha^{-1}, but considering the magnitude of values involved, the error is not relevant. The prediction equations, once again, don´t present any bias.

III.3. Fitting and evaluating the final system of equations

As explained in the methods section, each pair of prediction and projection models (allometric model and its difference formulation) was fitted together in order to assess and overcome possible violations of the regression assumptions and also to obtain initial values for the simultaneous fitting of the final system of link-functions.

The regression assumptions were guaranteed using the procedures already described in section 1. of the Results.

When simultaneously fitted, the parameter associated with N in the a parameter of the simple allometric model for G, becomes not significant, so only the k parameter is expanded with N.

In the multiple allometric model, the hdom prediction equation was fitted without the autocorrelation structure, but the growth equation was fitted with this structure. This resulted in a bias in the hdom prediction function, similar to the one found previously in the individual adjustments. The solution was to fix the values of both the a and k parameters in this equation (the ones where t was added as explanatory variable) to the ones obtained in the individual fitting, allowing only the parameter related to W_I to be estimated in the simultaneous fitting. This procedure corrected the bias without relevant decrease in the evaluation statistics.

The QQ plots (Figure 2) show that the final models have a close to normal distribution of the model errors, with some of them, like the volume functions, presenting a small deviation more evident in the tails. No apparent evidence of error heteroscedasticity was shown when analysing the studentized residuals of the final models.

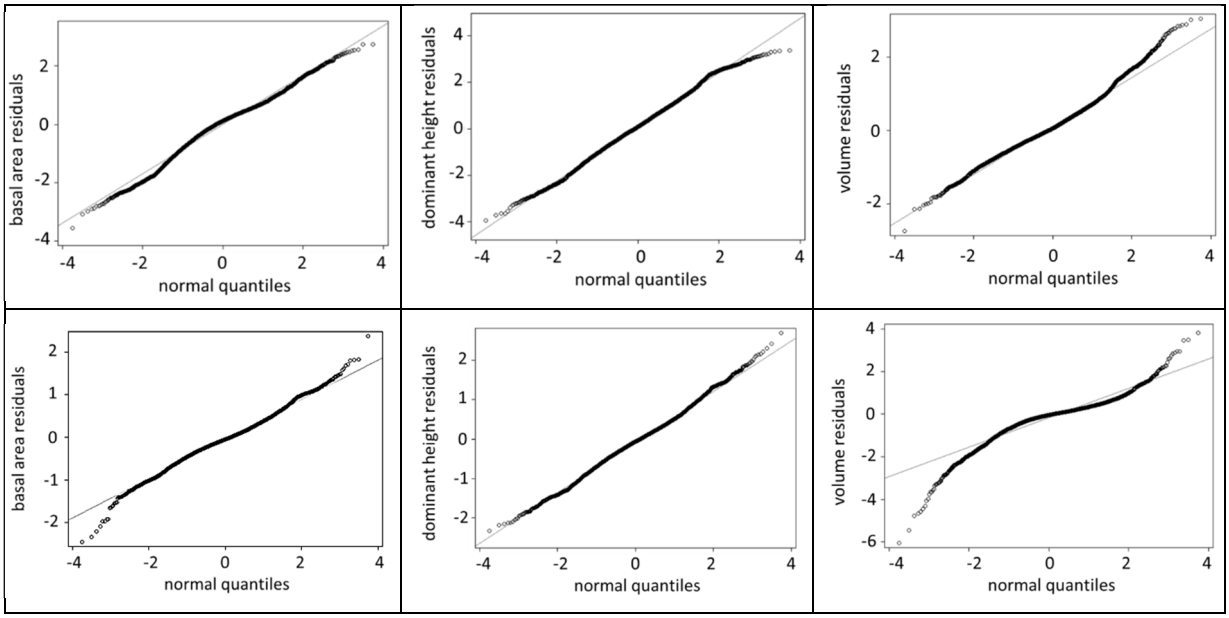


Figure 2. QQ plots for basal area (initialization on the left and projection on the right) in m^2 ha^{-1}, dominant height (initialization on the left and projection on the right) in m and volume (initialization on the left and projection on the right) in m^3 ha^{-1}.

The final equations obtained for the link functions, through a simultaneous fitting, can be seen in Tables 5 and 6.

Table 7 shows the statistics that characterize the prediction ability of the different allometric models that were selected. The values of the average PRESS residuals show that all the models have predictive ability with very small bias representing, for the 3 variables, less than 0.5% of the respective average value.

Table 5 – Final equations for the simple link-functions

Model	Analytical expression	

Basal area (G, m² ha^{-1})	(8) $G = k1 + k2\ \frac{N}{10000}\ W_a{}^{a}$ (Initialization)
	(9) $G_{t2} = G_{t1}\left(\frac{k1+k2\frac{N_2}{10000}\ W_{a\,2}{}^{a}}{k1+k2\frac{N_1}{10000}\ W_{a\,1}{}^{a}}\right)$ (Projection)
		k1 = 0. 741281;
		k2 = 0.808824;
		a = 0. 674882

Dominant height ($hdom$, m)	(10) $hdom = k1 + k2\ \frac{N}{10000}\ W_{wy}{}^{a1+a2\frac{N}{100000}}$ (Initialization)
	(11) $hdom_{t2} = hdom_{t1}\left(\frac{k1+k2\frac{N_2}{10000}\ W_{wy\,2}{}^{a1+a2\frac{N_2}{100000}}}{k1+k2\frac{N_1}{10000}\ W_{wy\,1}{}^{a1+a2\frac{N_1}{100000}}}\right)$ (Projection)
		k1= 5.232384;
		k2 = – 3.83914;
		a1 = 0.334543;
		a2 = 0.23738

Under bark volume (V_{u_st}, m² ha^{-1})	(12) $V_{u\_st} = k1 + k2\ \frac{N}{10000}\ W_{wy}{}^{a1+a2\frac{N}{100000}}$ (Initialization)
	(13) $V_{u\_st_{t2}} = V_{u\_st_{t1}}\left(\frac{k1+k2\frac{N2}{10000}W_{wy2}{}^{a1+a2\frac{N2}{100000}}}{k1+k2\frac{N1}{10000}W_{wy1}{}^{a1+a2\frac{N1}{100000}}}\right)$ (Projection)
		k1 = 1. 583779;
		k2 = 0. 841963;
		a1 = 0.977626;
		a2 = – 0.46796

Table 6 – Final equations for the multiple link-functions (continues)

Model	Analytical expression	

Basal area (G, m² ha^{-1})	(14) $G = k1 + k2\ \frac{N}{10000}\ W_a{}^{a1+a2\frac{N}{100000}}\ W_l{}^{b}$ (Initialization)
	(15) $G_{t2} = G_{t1}\left(\frac{k1+k2\frac{N_2}{10000}\ W_{a\,2}{}^{a1+a2\frac{N2}{100000}}\ W_{l\,2}{}^{b}}{k1+k2\frac{N_1}{10000}\ W_{a\,1}{}^{a1+a2\frac{N1}{100000}}\ W_{l\,1}{}^{b}}\right)$ (Projection)
		k1 = 0. 802248;
		k2 = 0.268406;
		a1 = 0.478916;
		a2 = 0.357292;
		b = 0.461176

Dominant height ($hdom$, m)	(16) $hdom = k1 + k2\,t\ W_{wy}{}^{a1+a2\frac{N}{100000}+a3\,t}\ W_l{}^{b}$ (Initialization)
	(17) $hdom_{t2} = hdom_{t1}\left(\frac{k1+k2\,t2\ W_{wy\,2}{}^{a1+a2\frac{N2}{100000}+a3\,t2}\ W_{l\,2}{}^{b}}{k1+k2\,t1\ W_{wy\,1}{}^{a1+a2\frac{N1}{100000}+a3\,t1}\ W_{l\,1}{}^{b}}\right)$ (Projection)
		k1= 3.90319;
		k2 = 0.084925;
		a1 = 0.641274;
		a2 = - 0.01809;
		a3= - 0.00572;
		b = - 0.57216

Table 6 – Final equations for the multiple link-functions (continuation)

Model	Analytical expression	
Under bark volume (V_{u_st}, m² ha^{-1})	(18) $V_{u\_st} = k1 + k2\ \frac{N}{10000}\ W_{wy}{}^{a1+a2\frac{N}{100000}}\ W_l{}^b$ (Initialization)	k1 = 1. 558135; k2 = 0. 766015; a1 = 0.974996; a2 = – 0.044424; b = 0.017652
	(19) $V_{u\_st}{}_{t2} = V_{u\_st}{}_{t1} \left(\frac{k1+k2\frac{N2}{10000} W_{wy2}{}^{a1+a2\frac{N2}{100000}}\ W_{l\,2}{}^b}{k1+k2\frac{N1}{10000} W_{wy1}{}^{a1+a2\frac{N1}{100000}}\ W_{l\,1}{}^b} \right)$ (Projection)	

Table 7 – Summary of the statistics used to characterize the predictive capacity of individual models, after the simultaneous adjustment

	Average	$R^2{}_{adj}$	SSR	MSR	$\widehat{r_p}$	$\widehat{\lvert r_p \rvert}$
simple equations						
Basal area initialization	14,79	0,9900	4434,2	0,6397	0,0064	0,5596
Basal area growth		0,9958	1999,9	0,2885	– 0,0342	0,3508
Dominant height initialization	18,71	0,9493	9658,0	1,3935	0,0894	0,8798
Dominant height growth		0,9802	3443,2	0,4968	– 0,0903	0,5032
Volume initialization	85,53	0,9999	4333,6	0,6252	0,1248	0,5051
Volume growth		0,9999	8544,5	1,2328	– 0,1570	0,4616
multiple equations						
Basal area initialization	14,79	0,9984	850,1	0,1227	– 0,0301	0,2127
Basal area growth		0,9987	966	0,1394	– 0,0586	0,1767
Dominant height initialization	18,71	0,9903	2676,7	0,3861	0,0979	0,3512
Dominant height growth		0,9962	1125,5	0,1624	– 0,0141	0,2011
Volume initialization	85,53	0,9999	4399,8	0.6348	0,1235	0,4827
Volume growth		0,9999	8231,7	1.1878	– 0.0999	0,4458

Basal area in m² ha^{-1}, dominant height in m and volume in m³ ha^{-1}.

The precision values are also very small, slightly higher for the prediction equations in relation to the corresponding projection equation, as expected. For the first the precision represents 4% of the average for G, 5% for hdom and less than 1% for V_{u_st}. Average precision (mean of the absolute value of the PRESS residuals) is, for all the models, below the measurement errors of the respective variables, indicating the good performance of the models.

The plots of the observed versus predicted values (not shown where) show a close to linear relationship with values close to the 1:1 reference line.

Discussion

The present research proposed the development of allometric models – for stand basal area (G), dominant height (hdom) and underbark volume without stump (V_{u_st}) – to link/hybridize the 3-PG and GLOBULUS models. Projection versions of the allometric models, for application in existing stands, were also fitted for each stand variable. The proposed methodology was successful as the allometric models provided very good fit to the data available. Stand basal area was best estimated as an allometric equation based on total aboveground biomass (W_a), crown biomass (W_c) or branches biomass (W_{br}) while the best regressor for underbark volume and dominant height was woody biomass (W_{wy}), stem biomass (W_s) or wood biomass (W_w). In a previous study, also for eucalypt in Portugal, Tomé *et al.* (2004) found similar results. This is an interesting result as it gives evidence of the relationship between basal area, related with the sapwood area, and the crown, in accordance with the pipe model theory (Shinozaki *et al.*, 1964a; 1964b, Valentine, 1985; Mäkelä, 1986).

Leaf biomass is generally not a good predictor for volume and dominant height which is understandable as leaf biomass levels out at canopy closure while underbark volume and dominant height are still increasing with considerable growth rates. However, if the allometric parameter is adjusted with age, prediction from leaf biomass improves considerably, although still being a poor prediction in relation to the ones obtained from other biomass components.

The analysis of the importance of calibrating the allometric constant and/or the respective scale parameter according to the characteristics of each particular stand showed the need to adjust at least one of them as a function of stand density. Note that, alternatively, stand density could have been used to adjust the allometric constant k with similar effect on the models predictive ability (results not shown). The signs of the parameters associated with stand density were positive in the allometric models for G and V_{u_st} but negative in the model for hdom. This implies that, for the same value of the biomass component used in the allometric model (W_a for G and W_{wy} for V_{u_st} and hdom), denser stands will have higher values of G and V_{u_st} but smaller

dominant height. These results show that stand density, by controlling individual tree allometry, impacts on the relationship between stand variables. Denser stands have trees with more cylindrical stems and smaller crowns leading to higher values of basal area and stem volume, not totally compensated by the increase in total aboveground biomass or woody biomass (for G and V_{u_st}, respectively). The data used in this research includes some spacing trials that confirm this finding (Figure 3).

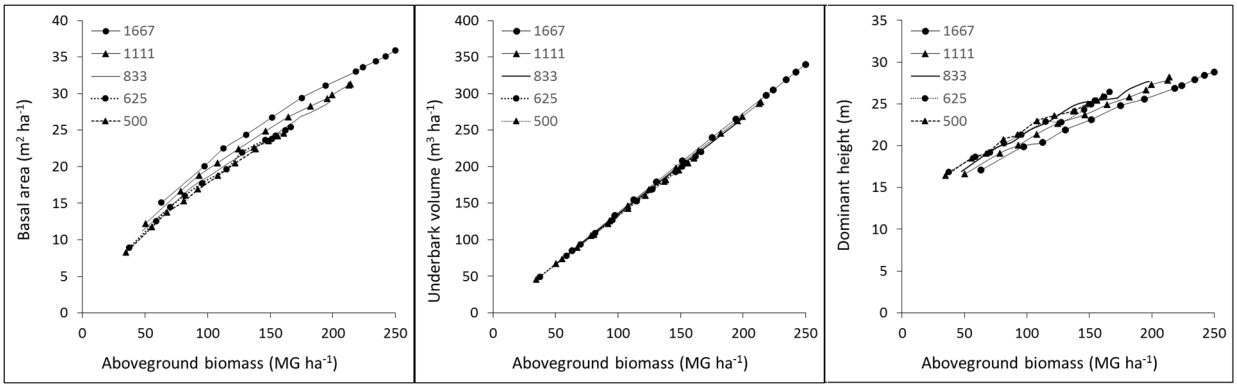


Figure 3. Spacing trial data - Relationship between the management oriented stand variables and aboveground biomass for different stand densities (1667, 111, 833, 625 and 500 trees ha^{-1}). The increase of stand basal area with stand density is clear as well as the decrease observed in dominant height. The positive relationship in underbark volume is not so clear due to the large range of the volume variable.

Other stand variables tested, namely stand age and a dummy variable for coppice, even if in some cases statistically significant, did not imply relevant decreases in the SSR values therefore they were not included in the proposed models in order to guarantee its simplicity, usually associated with robustness. Equations that do not include age as a regressor are always preferable in *Eucalyptus* stands (Tomé *et al.*, 2006a) as stand age is very often not available, for instance in National Forest Inventories, as annual growth rings are not identifiable in eucalypt wood.

The models proposed represent an important improvement of the information to managers provided by the 3-PG model. With these equations, the variables that were already part of the 3-PG output, stand basal area and underbark volume, are predicted with more precision and less bias, namely when the model is used to project existing stands. The addition of an equation to predict dominant height allows the use of all the equations from the GLOBULUS model, namely the system of equations to predicted merchantable volumes, with or without bark, to any top diameter or the diameter distribution model from Mateus and Tomé (2011). The models just need stand biomass per tree component and stand density as input, therefore

they can be used to complement the output of any process based model that provides prediction of total aboveground and woody biomass over time. It is clear that the proposed hybridization has the potential to deliver robust management tools that can be used with confidence under changing environments. The use of a process based model helps to overcome limitations of empirical based models, for instance by allowing the prediction of potential productivity in areas where there is no previous information and by making the models sensitive to changes in weather conditions and silvicultural practices. However, one must not forget that the hybridization results in growth projections subject to the assumptions of both models.

The analysis of the statistics obtained after the simultaneous fitting (Table 4) shows that overall the equations have good prediction ability, with high values of adjusted R^2_{adj}, small bias and high precision.

Conclusions

This study aimed at developing a methodology for the hybridization of whole stand process based models so that the resulting model is responsive to climate and silviculture changes but still provides all the information that is required for forest management decisions. The system of equations presented on Tables 5 and 6 are the most important conclusion of the study, an important added value to the 3-PG growth module, allowing this model to provide diversified information for management, also applicable to other process based models.

The development of the system of equations brought some other conclusions that advance knowledge about the stand level allometric relationships:

- Stand variables that are important for forest management (basal area, underbark volume and dominant height) can be accurately predicted (low bias and high precision) from the allometric relationship with some component of stand biomass. Stand basal area is better predicted from total aboveground biomass while prediction of underbark volume and dominant height is better if using woody biomass as predictor.
- Stand level allometric relationships with total aboveground biomass or woody biomass must be calibrated with stand density by expanding at least one of the allometric parameters as a linear function of stand density. This adjustment implies an increase of the parameter with stand density for basal area and volume but a decrease in the relationship for dominant height.
- Leaf biomass is not a good predictor of the management oriented stand variables, due to the fact that it levels out after canopy closure while other stand variables are still in the growing phase.

As a final conclusion one may say that the establishment of good links – here denoted as link-functions and using stand level allometric relationships – between a process based model and a traditional growth and yield model is a good approach when the objective is to obtain a model with the ability to predict potential productivity, sensitive to the impact of climate change on growth, able to predict the impact of intensive silvicultural treatments such as irrigation and fertilization.

Acknowledgments

This paper is a part of the PhD project of the first author, which is funded by a scholarship (SFRH/ BD/ 46594/ 2008) granted by the Portuguese Science Foundation.

Permission to use data from the ALTRI and PORTUCEL Portuguese pulp and paper companies is also greatly acknowledged.

References

Almeida, A.C., Maestri, R., Landsberg, J.J., Scolforo, J.R.S. 2003. Linking process-based and empirical forest models in Eucalyptus plantations in Brazil. In: Amaro, A., Reed, D., Soares, P. (Eds), Modelling Forest Systems. CAB International. pp. 63-74.

António, N., Tomé, M., Tomé, J., Soares, P., Fontes, L. 2007. Effect of tree, stand and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research 37: 895-906.

Baldwin, V.C. Jr, Burkhart, H. E., Westfall, J. A., Peterson, K. D. 2001. Linking growth and yield and process models to estimate impact of environmental changes on growth of loblolly pine. Forest Science. 47(1): 77-82.

Battaglia, M., Sands, P. J., Candy, S. G. 1999. Hybrid growth model to predict height and volume growth in young *Eucalyptus globulus* plantations. Forest Ecology and Management. 120: 193-201. DOI: 10.1016/S0378-1127(98)00548-9

Bravo-Oviedo, A., Tomé, M., Bravo, F., Montero, G., Rio, MD. 2008. Dominant height growth equations including site atributes in the generalized algebraic difference approach. Canadian Journal of Forest Research. 38: 2348-2358.

Bugmann, H. 2001. A review of forest gap models. Climatic Change 51(3-4): 259-305.

Burkhart, H, Farrar, K.D., Amateis, R.L., Daniels, R. R. 1987. Simulation of individual tree growth and stand development in loblolly pine plantations on cutover, site-prepared areas. Publication No. FWS-1-87, Department of Forestry, Virginia Tech, Blacksburg, Virginia.

Burkhart, H. E., Tomé. M. 2012. Modelling Forest Trees and Stands. Springer. Dordrecht, Netherlands.

Fontes, L., Landsberg, J., Tomé, J. Tomé, M., Pacheco, C. A., Soares, P., Araújo, C. 2006. Calibration and testing of a generalized process-based model for use in Portuguese Eucalyptus plantations. Canadian Journal of Forest Research 36(12): 3209-3221.

Kimmins, J.P., Blanco, J.A., Seely, B., Welham, C., Scoullar, K. 2010. Forecasting forest futures: a hybrid modelling approach to the assessment of sustainability of forest ecosystems and their values. Earthscan, London.

Kimmins, J.P., Mailly, D., Seely, B. 1999. Modelling forest ecosystem net primary production: the hybrid imulation approach used in FORECAST. Ecological Modelling. 122: 195–224.

Landsberg, J. J., Waring, R. H. 1997. A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and Management. 95: 209-228.

Landsberg, J.J. 1986. Physiological ecology of forest production. Academic Press, New York.

Landsberg, J.J., Waring, R.H., Coops, N.C., 2003 Performance of the forest productivity model 3-PG applied to a wide range of forest types. Forest Ecology and Management. 172: 199–214.

Levy, P.E., Hale, S.E., and Nicoll, B.C. 2004. Biomass expansion factors and root:shoot ratios for coniferous tree species in Great Britain. Forestry, 77(5): 421–430.

Luxmoore R.J., Hargrove W.W., Tharp M.L., Mac PostW., Berry M.W., Minser K.S., Cropper W.P., Johnson D.W., Zeide B., Amateis R.L., Burkhart H.E., Baldwin V.C., Peterson K.D. 2002. Addressing multi-use issues in sustainable forest management with signal-transfer modelling. Forest Ecology and Management. 165(1-3): 295-304.

Mäkelä, A. 1986. Implication of the pipe model theory on dry matter partitioning and height growth in trees. Journal of Theoretical Biology. 123:103 - 120

Mäkelä, A. 1997. A carbon balance model of growth and self-pruning in trees based on structural relationships. Forest Science 43(1), 7–24.

Mäkelä, A. 2002. Derivation of stem taper from the pipe theory in a carbon balance framework. Tree Physiol. 22: 891–905.

Mäkelä, A. 2009. Hybrid Models of Forest Stand Growth and Production. In: Dykstra, D.P.; Monserud, R.A., tech.eds. 2009. Forest growth and timber quality: Crown models and simulation methods for sustainable forest management. Proceedings of an international

conference. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. pp 43-47.

Mäkelä, A., Landsberg, J., Ek, A.R., Burk, T.E., Ter Mikaelian, M., Agren, G.I., Oliver, C.D., and Puttonen, P. 2000. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. Tree Physiol. 20(5-6): 289-298. DOI: 10.1093/treephys/20.5-6.289.

Mäkelä, A., Mäkinen, H. 2003. Generating 3D sawlogs with a process-based growth model. Forest Ecology and Management. 184: 337-354.

Matala, J., Ojansuu R., Peltola H., Sievanen R., Kellomäki S. 2005. Introducing effects of temperature and CO2 elevation on tree growth into a statistical growth and yield model. Ecol Modell 181(2-3), 173-190.

Matala, J., Ojansuu, R., Peltola, H., Raitio, H., Kellomäki, S. 2006. Modelling the response of tree growth to temperature and CO2 elevation as related to the fertility and current temperature sum of a site. Ecological Modelling. 199(1): 39-52.

Mateus, A., Tomé, M. 2011. Modelling the diameter distribution of eucalyptus plantations with Johnson's probability density function: parameters recovery from a compatible system of equations to predict stand variables. Annals of Forest Science. 68(2): 325-335.

Mohren, G.M.J., Burkhart, H.E. 1994. Contrasts between biologically-based process models and management-oriented growth and yield models. Forest Ecology and Management. 69:1–5.

Myers, R. H. 1986. Classical and Modern Regression with Applications. Second edition. Duxbury Classic Series.

Nunes, L., Patrício, M., Tomé, J., Tomé, M. 2011. Modelling dominant height growth of maritime pine in Portugal using GADA methodology with parameters depending on soil and climate variables. Annals of Forest Science. 68(2): 311-323.

Oliveira, T., Tomé, M. (2015). A system of compatible models to predict and project crown ratio and crown length for eucalypt plantations. Submitted.

Peng, C., Liu, J., Dang, Q., Apps, MJ., Jiang, H. 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics, Ecological Modelling. 153(1–2): 109-130.

Pinjuv, G.L., Mason, E.G., Watt, M. 2006. Quantitative validation and comparison of a range of forest growth model types. Forest Ecology and Management. 236: 37-46.

Robinson, A. P., Ek, A. R. 2003. Description and validation of a hybrid model of forest growth and stand dynamics for the Great Lakes region. Ecological Modelling. 170(1): 73-104.

Sands, P. J., Landsberg, J. J. 2002. Parameterisation of 3-PG for plantation grown *Eucalyptus globulus*. Forest Ecology and Management. 163: 273–292.

SAS Institute Inc. 2009a. SAS/ETS ® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 2876 p.

SAS Institute Inc. 2009b. SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC: SAS Institute Inc. 7886 p.

SAS Institute Inc. 2009c. SAS ® 9.2 Output Delivery System: User's Guide. Cary, NC: SAS Institute Inc. 984 p.

Schabenberger, O., Pierce, F.J. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press LLC 2000. Boca Raton, FL. pp. 211-213 and pp. 539-545.

Shinozaki K., Yoda K., Hozumi K., Kira T. 1964a. A quantitative analysis of plant form-the pipe model theory. I. Basic analyses. Japanese Journal of Ecology 14: 97–105.

Shinozaki K., Yoda K., Hozumi K., Kira T.1964b. A quantitative analysis of plant form- the pipe model theory: II. Further evidence of the theory and its application in forest ecology. Japanese Journal of Ecology. 14: 133–139.

Tomé, J., Tomé, M., Barreiro, S., Paulo, J. A. 2006a. Age-independent difference equations for modelling tree and stand growth. Canadian Journal of Forest Research, 36: 1621-1630.

Tomé, M., Faias, S. P., Tomé, J., Cortiçada, A., Soares, P., Araújo, C. 2004. Hybridizing a stand level process-based model with growth and yield models for *Eucalyptus globulus* plantations in Portugal. In: Borralho, N. M. G., Pereira, J. S., Marques, C., Coutinho, J., Madeira, M. & Tomé, M. (ed.), Eucalyptus in a changing world. Proc. Iufro Conf., Aveiro, 11-15 Oct. (RAIZ, Instituto de Investigação da Floresta e do Papel, Portugal), pp.290-297.

Tomé, M., Oliveira, T., Soares, P. 2006b O modelo GLOBULUS 3.0. Publicações GIMREF - RC2/2006. Universidade Técnica de Lisboa. Instituto Superior de Agronomia. Centro de Estudos Florestais. Lisboa. http://hdl.handle.net/10400.5/1760

Tomé, M., Ribeiro, F., Faias, S. 2007a. Relação hipsométrica geral para *Eucalyptus globulus* Labill. em Portugal. Silva Lusitana 15(1): 41-56.

Tomé, M., Ribeiro, F., Soares,P. 2001. O Modelo GLOBULUS 2.1. Relatórios Técnico–Científicos do GIMREF, nº 1/2001, Departamento Engenharia Florestal, ISA, Lisboa.

Tomé, M., Tomé, J., Ribeiro, F., Faias, S. 2007b. Equações de volume total, volume percentual e de perfil do tronco para *Eucalyptus globulus* Labill. em Portugal. Silva Lusitana 15(1): 25-40.

Turnbull, C.R.A., McLeod, D.E., Beadle, C.L., Ratkowsky, D.A., Mummery, D.C., Bird, T. 1993. Comparative early growth of Eucalyptus species of the subgenera Monocalyptus and Symphyomyrtus in intensively-managed plantations in southern Tasmania. Australian Forestry 56(3): 276-286.

Valentine, H.T. 1985. Tree-growth models: derivations employing the pipe-model theory. Journal of Theoretical Biology. 117:579–585

Wang, Y.P., Jarvis, P.G. 1990. Description and validation of an array model - Maestro. Agricultural and Forest Meteorology. 51: 257-280.

Appendices

Table A.1. – Residual sum of squares (SSR) and R^2 values for the link functions for *G*, V_{u_st} and *hdom* using alternative expansions for the *a* parameter.

	Wa		Wwy		Ws		Ww		Wl		Wc		Wbr	
							G							
base	20020,2		24218,62		27015,5		27557,5		84602		24239,2		7176,8	
N	13744,9	+	16436,7	+	18699,3	+	20098	+	83388,1	-	24177,3	-	6428,7	+
t	19906,3	-	24168,6	-	27004,2	-	27554,5	+	36964,8	+	11768,4	+	4695,6	+
coppice	20020	-	24212,6	-	26999,6	-	27942,3	-	84458,2	+	23985,4	+	6917,9	+
N,t	13737,4	+,-	16436,6	+,+	18683,6	+,+	20017,4	+,+	36949,5	-,+	11667,3	+,+	3624,2	+,+
N,coppice	13519,5	+,-	16088,5	+,-	18283,8	+,-	19541,9	+,-	82995,9	-,+	23855,9	-,+	6299,9	+,+
t,coppice	19903,8	-,-	24158,1	-,-	26985	-,-	27491,6	+,-	36225,4	+,+	11179,3	+,+	4274	+,+
Full model	13504	+,-,-	16087	+,-,-	18279,2	+,+,-	19495,4	+,+,-	36120,7	-,+,+	11152,5	+,+,+	3398,9	+,+,+
R^2 base	0,9934		0,9920		0,9910		0,9909		0,9720		0,9920		0,9976	
R^2 full	0,9955		0,9947		0,9939		0,9935		0,9880		0,9963		0,9989	
							$V_{u\ st}$							
base	65822		57371,5		60169,2		42137,3		18557341		7971653		3278674	
N	57650,6	+	29055,5	+	24338,9	+	25012,1	+	1,8E+08	-	7565619	-	3179999	-
t	57378	-	51781,6	-	58036,1	-	42132,7	+	10288967	+	5487977	+	2758285	+
coppice	59617,3	+	53954,7	+	58294,2	+	42130,6	+	18546859	+	7937421	+	3230280	+
N,t	50252,3	+,-	24686,7	+,-	23004,8	+,-	24980,2	+,+	10090707	-,+	5297800	-,+	2690859	-,+
N,coppice	53571,2	+,+	28195,2	+,+	24219,4	+,+	24606,8	+,-	17958442	-,+	7458133	-,+	3100165	-,+
t,coppice	52393,5	-,+	49134,8	-,+	56521,3	-,+	42127,1	-,+	10208942	+,+	5402303	+,+	2679652	+,+
Full model	46992,8	+,-,+	24128,3	+,-,+	22945,8	+,-,+	24592,4	+,+ -	9926862	-,+,+	5139073	-,+,+	2579261	-,+,+
R^2 base	0,9996		0,9996		0,9996		0,9997		0,8841		0,9502		0,9795	
R^2 full	0,9997		0,9998		0,9999		0,9998		0,9783		0,9679		0,9839	
							hdom							
base	43810,9		39320		36488		35785,3		160454		120215		91873,9	
N	36212,4	-	32787,5	-	30389,3	-	29248,1	-	148779	-	107949	-	83174	-
t	42588,5	+	38398,9	+	35692,3	+	34864,4	+	119907	+	97314,7	+	81954,8	+
coppice	43213,7	+	38768,7	+	35978,9	+	35409,9	+	160052	+	119604	+	91186,6	+
N,t	35508,4	-,+	32219,2	-,+	29882,4	-,+	28648,6	-,+	112087	-,+	88662	-,+	74308,6	-, +
N,coppice	34143,9	-,+	30946,8	-,+	28704,7	-,+	27777,4	-,+	146476	-,+	105084	-,+	80818,6	-, +
t,coppice	41744,8	+,+	37638,4	+,+	34994,4	+,+	34307,6	+,+	117956	+,+	95503,7	+,+	80517,5	+,+
Full model	33125,1	-,+,+	30096,3	-,+,+	27934,7	-,+,+	26903,4	-,+,+	107404	-,+,+	84131,1	-,+,+	70801,9	-,+,+
R^2 base	0,9911		0,9921		0,9926		0,9928		0,9676		0,9757		0,9814	
R^2 full	0,9933		0,9939		0,9944		0,9946		0,9783		0,9830		0,9857	

Notes:

- the columns with signs indicate the signs of the coefficients associated with the variables included in the expansion of the *a* parameter
- the base model refers to the formulation $G/V_{u\_st}/hdom = k\ Wi^{a}$
- the full model used N, t and coppice in the expansion of the *a* parameter
- the shaded values of SSR indicate that one of the variables is not significant

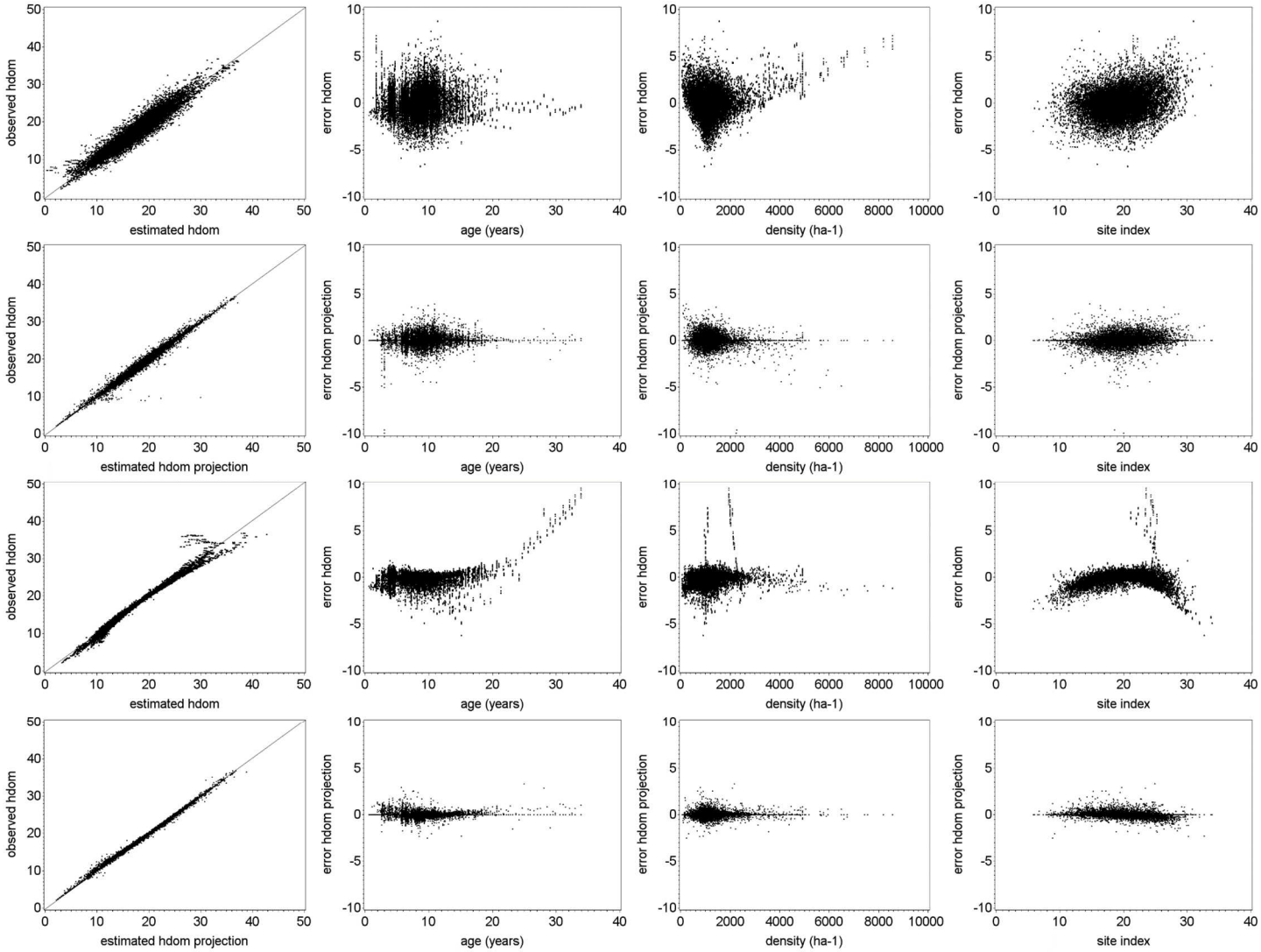


Figure A.1. Validation of the prediction and projection of dominant height (m) with the link functions proposed - simple equations above and multiple allometric equations below.

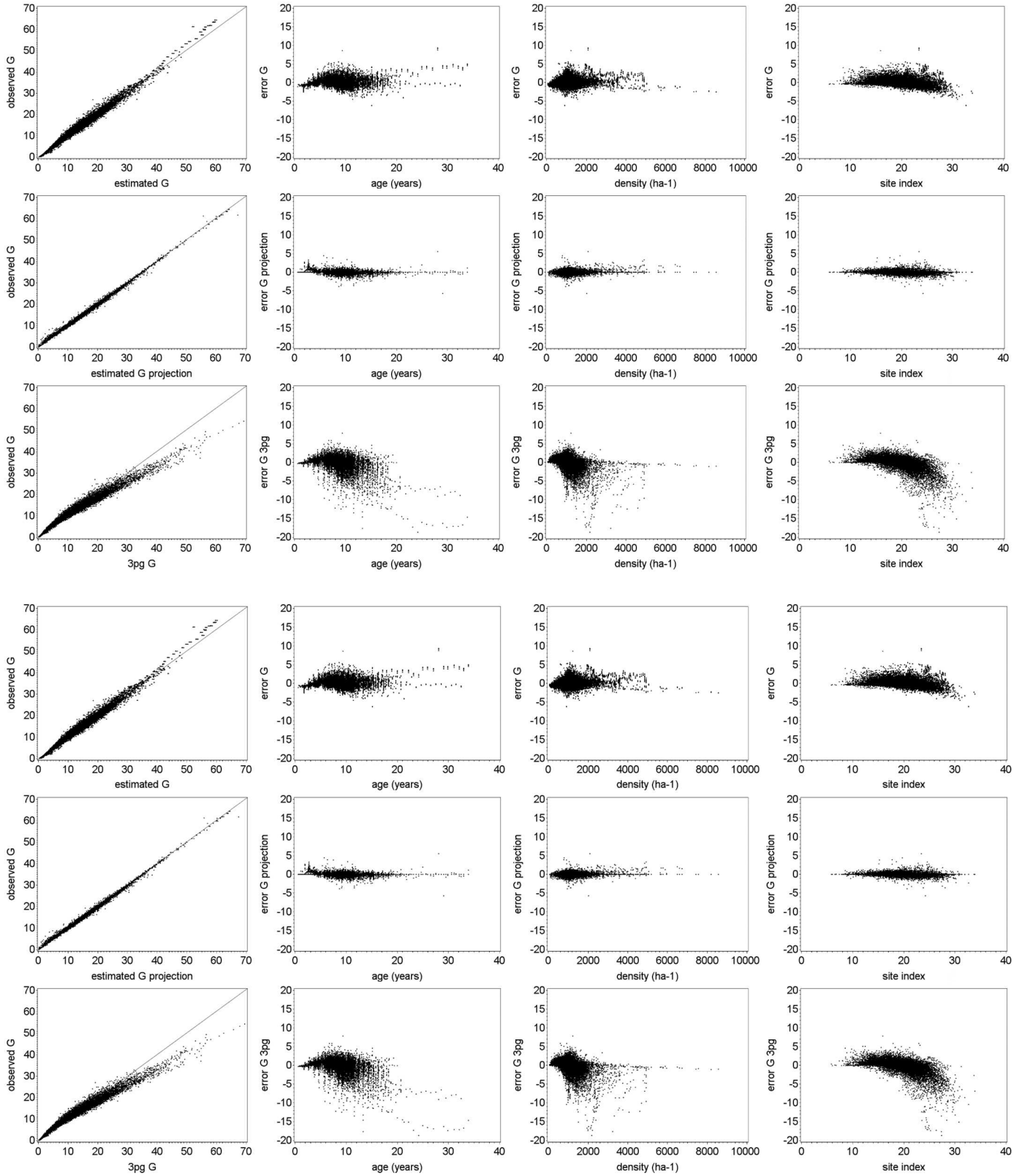


Figure A.2. Validation of the prediction of basal area (m^2 ha^{-1}) with the link functions proposed (prediction and projection of G with the simple equations in 1st and 2nd rows, respectively and prediction and projection of G with the multiple allometric equations in 4th and 5th rows, respectively) and according to the 3-PG model (3rd and 6th rows).

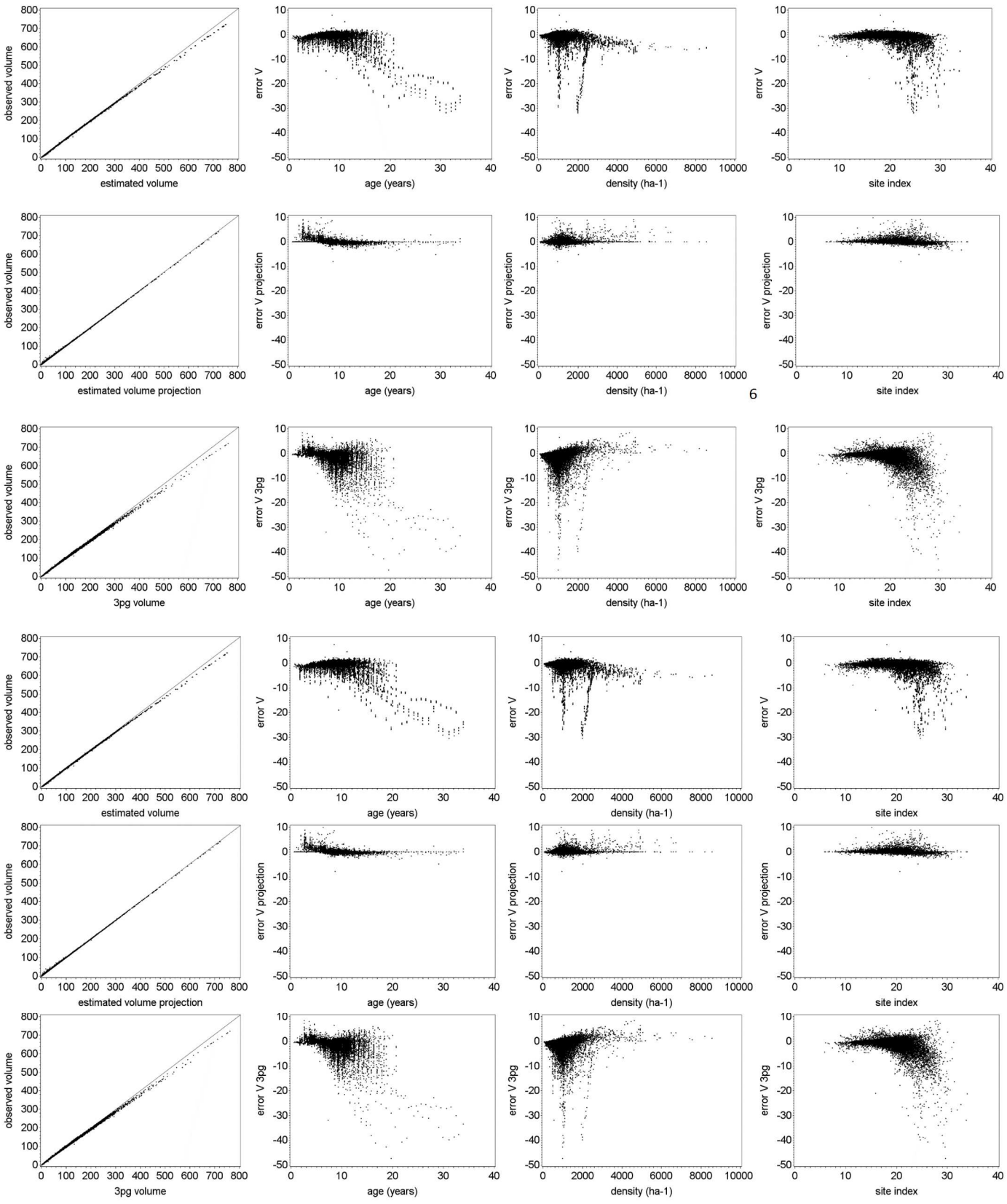


Figure A.3. Validation of the prediction of volume (m^3 ha^{-1}) with the link functions proposed (prediction and projection of volume with the simple equations in 1st and 2nd rows, respectively and prediction and projection of volume with the multiple allometric equations in 4th and 5th rows, respectively) and according to the 3-PG model (3rd and 6th rows).

Final remarks

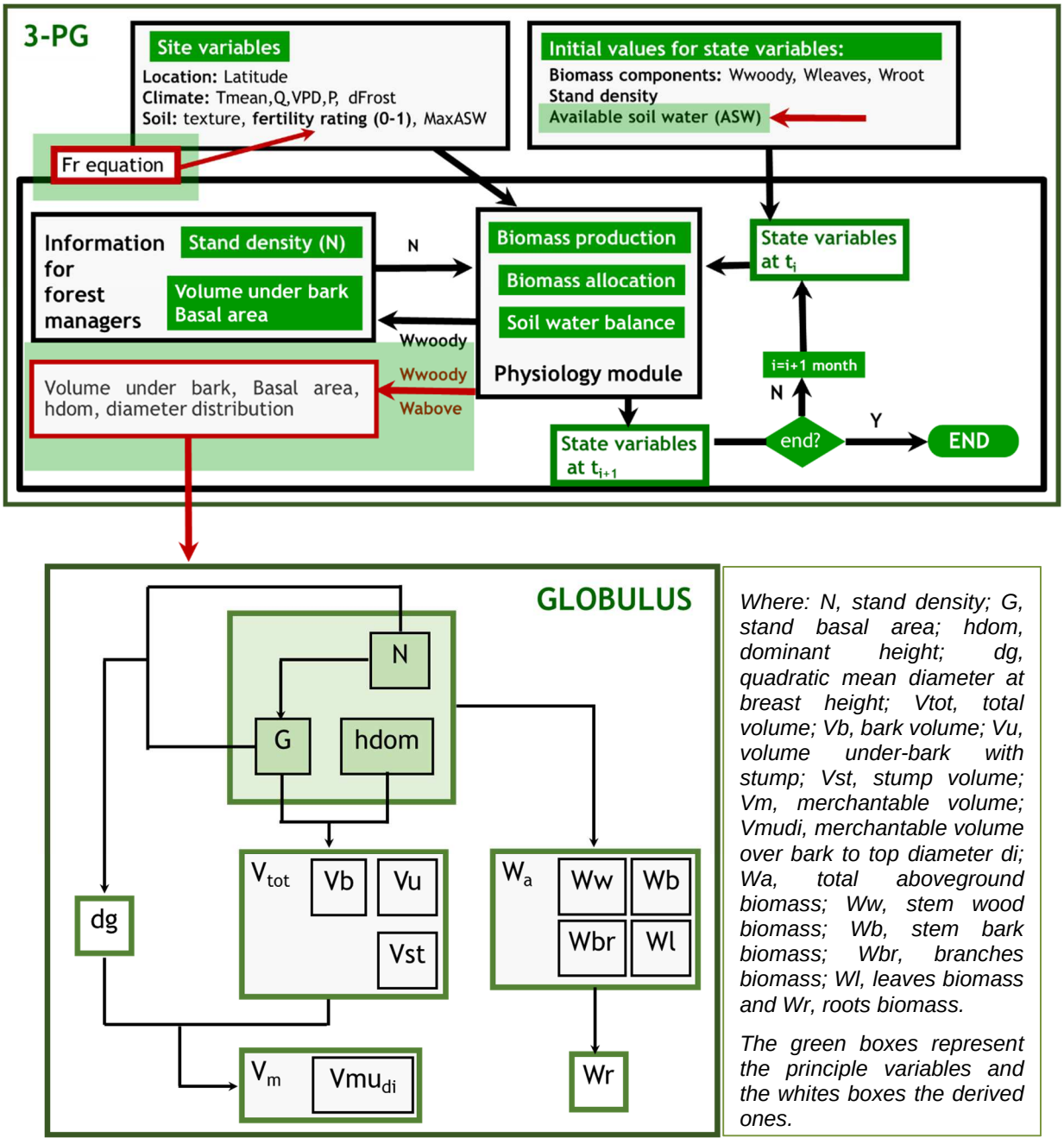
Forest management is facing multiple new challenges and has been walking away from traditional statistical yield and growth models to embrace process-based and hybrid models to be able to face new demands. Silviculture has to account not only multiple objectives, but also new ones, which brings the need for new tools. But there is no sense in developing models if they are not to be used, so models must be useful and provide information to forest managers and decision makers (Vanclay, 1994). This does not mean that all results must be directly applied. Results can also be used as guidelines to explore alternatives, evaluate consequences of specific actions or evaluate the system's sensitivity to a certain disturbances (Landsberg, 2003).

Modern forestry has been asked for complex responses that can only be provided by elaborated computerized tools (Pretzsch *et al.*, 2002). Process-based models (PBMs) are very appealing, not only from a scientific point of view, but also from a management perspective. Hybrid models combine statistical and PBMs to take advantage of the best of both type of models and try to address the best way possible the limitations of them both. There are several approaches used to hybridize models and the degree of improvement obtained is very variable, but hybrid models allow statistical equations to be used in new situations and offer a new range of outputs that are useful for many different purposes. They stand between statistical and PBMs, blurring the lines between them and allowing a whole new potential of modelling frameworks. As long as the knowledge and understanding of physiological processes and the complexity of questions that are presented to modern forestry/forest management, the use and development of hybrid models will continue to increase.

The purpose of this work was to obtain a simple model that can be used on current forest management, because it gives useful information, and that can account for climate changes and management alternatives such as fertilization and irrigation. All the work that was developed had as final goal the improvement of the 3-PG (a process based model that uses climate data and that can be used to simulate irrigation and changes in fertilization) usability, so an equation to predict the FR parameter from site characteristics was developed and the outputs of the model were enhanced and complemented. One complementary result from this work was the validation of the method to predict the maximum available soil water, one other parameter of the 3-PG model.

Figure 1 translates the work in a graphical way. It represents the 3-PG model with the improvements developed in this thesis in red. The outputs of the model using the link-functions give better predictions of the variables, but serve also as inputs to the GLOBULUS model and

allow the estimation of other variables such as the merchantable volumes and the other components of biomass.



Where: N, stand density; G, stand basal area; hdom, dominant height; dg, quadratic mean diameter at breast height; Vtot, total volume; Vb, bark volume; Vu, volume under-bark with stump; Vst, stump volume; Vm, merchantable volume; Vmudi, merchantable volume over bark to top diameter di; Wa, total aboveground biomass; Ww, stem wood biomass; Wb, stem bark biomass; Wbr, branches biomass; Wl, leaves biomass and Wr, roots biomass.

The green boxes represent the principle variables and the whites boxes the derived ones.

Figure 1. The improved hybrid model.

The first chapter of the thesis describes the development of two systems of equations to predict, in a compatible way, crown ratio and crown length and to project crown length in *Eucalyptus globulus* plantations in Portugal. Both systems of equations consider the specific characteristics of the stand as regressors, but they differ by considering or not stand age as

one of the regressors. Even in eucalyptus stands, stand age is not always available, so a system that doesn´t take it in consideration is of added value.

Crown dimensions can be estimated using a Richards function that has as parameters stand density, dominant height, diameter at breast height and age, if available.

The equations to predict crown size are important not only due to their importance in growth and yield models and as variables in crown biomass equations (António *et al*., 2007), but also in forest fires, for example. Crown dimension influences the size and the spread of forest fires, so this is valuable information.

The second chapter describes the development of two systems of equations to predict stand level total aboveground biomass and biomass per tree component (stem, bark, branches and leaves). To allow a simple use of the equations, only stand variables that usually come as result of forest inventories are used as regressors. The equations are precise and with a high predictive ability considering only as variables stand age, stand density, dominant height, basal area and cultural regime.

Having the crown dimensions equations and the biomass equations, the database that was needed to the next part of the work was completed.

The third chapter describes the improvement of the output of the 3-PG model by complementing it with the diameter distribution model. Equations where developed for minimum, average, median and maximum diameter. The equations where built in a way to obtain estimates that observe logical relationships according to biological principles. Using the diameter variables predicted with this equations and the stand variables basal area, dominant height, stand density and quadratic mean diameter, the Visual Basic program developed by Marto *et al.* (2009) that is implemented in the sIMFLOR interface (Faias *et al.*, 2012) can be used to estimate the parameters of a Johnson's SB distribution. This parameters can be then used to estimate the diameter distribution for any stand.

The fourth chapter describes the development of a generalized model to predict FR from soil characteristics. The site-specific models can predict FR values from soil variables like the water content, percentage of sand, silt and clay and nutrients (e.g. K, Ca, Na, Mg, among others) with values of R^2 between 0.55 and 0.97. Considering the previous works of Stape *et al.* (2004) and Vega-Nieva *et al.* (2013), this was the expected result. The global model did not perform that well, with a value of R^2 of only 0.34 for the model that considered al the fertilization treatments and 0.58 for the model considering only the traditional fertilization plots. The differences between the results of the models developed for Spain and the ones resulting from this work might be related to soil differences or the fact that in this work the fertilization component was also considered. The results don´t allow a definitive answer, but establish a

starting point for future research. Some other ways to look at this problem would be to complement the data with foliar analysis to see if the fertilization is actually being used in the plant growth, to collet new data from other sites, to try other modelling approaches such as cluster analysis or decision trees or even try to use a function bounded between 0 and 1 and not a linear one to model FR.

Asides from the FR modelling, some other results were found in this work, namely the sensitivity of the 3-PG model to the initial values of biomass - that are even greater than the model sensitivity to available soil water – and the validation of the use of the Domingo Santos *et al.* (2006) expression to estimate the available soil water in Portugal. Figure 2 plots the values of maximum available soil water (ASW) estimated with the method Domingo Santos *et al*. (2006) versus the values determined in the field by Professor Arruda when the soil pits where open. The differences between them are very small and there is only one site where the values given by Professor Arruda are much more conservatives that the ones estimated by the expression, but values of ASW over 300mm are not that usual and the real values of ASW for those sites might be between the values estimated by the expression and the ones determined in the field. In a global way, the method of Domingo Santos *et al*. (2006) shoes good results and can be used in Portugal.

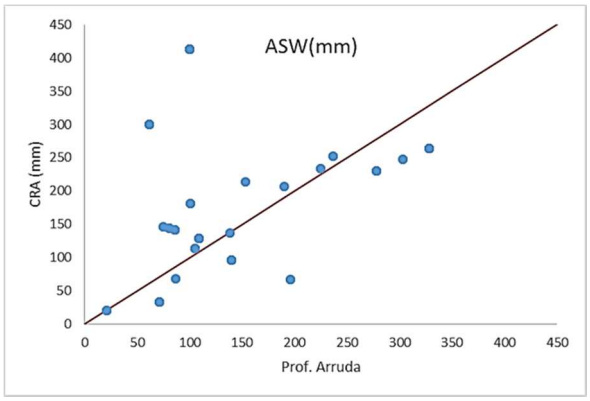


Figure 2. Values of the maximum available soil water estimated with the method of Domingo Santos et al. (2006) versus the values determined in the field by Professor Arruda.

The last chapter describes the main and last of the work, the development of a hybridization methodology between the 3-PG and the GLOBULUS models.

The development of the system of equations that predict volume, basal area and dominant height from the biomass estimates of 3-PG also allow the connection to the GLOBULUS model, that allows the estimation of other operationally relevant variables. Basal area, underbark volume and dominant height can be predicted with low bias and high precision from allometric

relationships with total aboveground biomass (basal area) or woody biomass (underbark volume and dominant height) and using other variables like stand density and leaf biomass and stand age, when available. Figure 3 plots the application of the hybridization equations for volume and basal area to data from one of the sites used in the FR work and shows the differences between the real values of volume and basal area and the ones estimated using 3-PG and the link-functions developed. 3-PG tends to overestimate both values while the equations developed give better results, with values closer to the real ones.

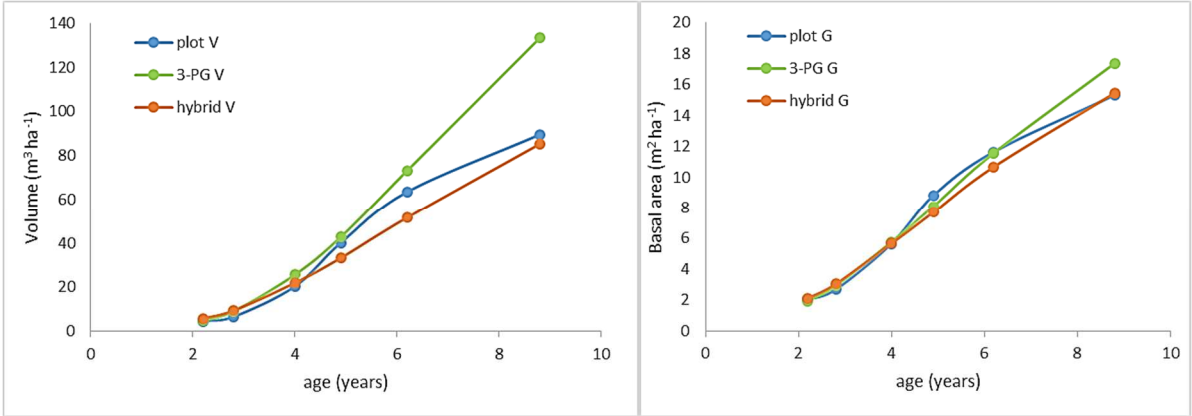


Figure 3. Plots of values of volume and basal area: the real values, the ones estimated by the 3-PG model and the ones estimated with the link-functions developed.

The establishment of good links between a process based model and a traditional growth and yield model is a good approach when the objective is to obtain a model with the ability to predict potential productivity, sensitive to the impact of climate change on growth, able to predict the impact of intensive silvicultural treatments such as irrigation (3-PG can be used to simulate irrigation) and fertilization.

Where will forest models go to in the future is hard to predict. New challenges will continue to be presented to forest management and the need for new and different models will continue. As noted by Furnival (1987), the key to making future progress will be avoiding the mistakes of the past, which can come in many forms. Future scenarios are uncertain, and considering that, one might ask how precise do model predictions have to be to be useful? No matter what the future holds, models will still have to be developed using the existing knowledge, have to be rigorously tested and made accessible to a wide array of users. These principles will not likely change with time or modelling approach. Neither will the base of the modelling process, data. Some modellers say that it might take years before they can gather the ideal data to develop a model (Vanclay, 1994). Gathering data is a hard and time consuming task. The available information in, most the times, lacking the information needed to develop a model.

The way inventories are designed does not always have in mind the need for accurate data on growth, so this is always be the first step in modelling.

Obtaining representative data of eucalyptus stands in Portugal was not a difficult task due to the long-time collaboration between the university and the pulp and paper industry and the existence of several permanent plots and trials, but quantity does not mean quality and some specific data may not even be available. This was the case of soil related information. The available soil maps are too general to be used, so field data had to be collected. This type of data is expensive and time consuming, so in the end, the idea persisted that the amount of information collected was too small. Soil characteristics are too variable in space and numerous interactions exist between soil components, water and nutrients that are difficult to account for.

References

António, N., Tomé, M., Tomé, J., Soares, P., Fontes, L. (2007). Effect of tree, stand and site variables on the allometry of *Eucalyptus globulus* tree biomass. Canadian Journal of Forest Research 37: 895-906.

Domingo Santos, J.M., Fernández de Villarán, R., Corral, E., Rapp, I. 2006. Estimación de la capacidad de retención de agua en el suelo: revisión del parámetro CRA. Invest Agraria: Sistemas y Recursos Forestales 15:14 –23.

Faias, S.P., Palma, J.H.N., Barreiro, S.M., Paulo, J.A., Tomé, M. 2012. Resource communication. sIMfLOR – Platform for the Portuguese forest simulators. Forest Systems 21(3): 543-548.

Furnival, G.M. 1987. Growth and yield prediction: Some criticisms and suggestions. In Chappell, H.N., Maguire, D.A., (Eds.), Predicting forest growth and yield: Current issues, future prospects. Contribution Number 58. Institute of Forest Resources, University of Washington, Seattle, WA, pp 22-30.

Landsberg, J.J. 2003. Physiology in forest models: history and the future. Forest Biometry Modelling and Information Sciences, volume 1: 49-63.

Marto, M., Palma, J., Mateus, A., Tomé, M. 2009. Computer program for estimation of Johnson´s S_B parameters using a parameter recovery method. Publicações Científicas Forchange PC-X/2009. Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa.

Pretzsch, H., Biber, J., Ďurský, K., Gadow, K. Von, Hasenauer, H., Kändler, G., Kenk, G., Kublin, E., Nagel, J., Pukkala, T., Skovsgaard, J. P., Sodtke, R., and Sterba, H. 2002. Recommendations for Standardized Documentation and Further Development of Forest Growth Simulators. Forstw. Cbl, 121: 138-151.

Vanclay, J.K. 1994. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests. CAB International, Wallingford, UK.

Vega-Nieva, D.J., Tomé, M., Tomé, J., Fontes, L., Soares, P., Ortiz, L., Basurco, F., Rodrígez-Soalleiro, R. 2013. Developing a general method for the estimation of the fertility rating parameter of the 3-PG model: application in *Eucalyptus globulus* plantations in northwestern Spain. Canadian Journal of Forest Research, 43: 627-636.